Improved models for the prediction of breeding values in trees

by

Gregory Wojtek Dutkowski B. F. Sc. (Hons) (University of Melbourne)

A thesis submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

University of Tasmania

(July, 2005)

Declarations

This thesis contains no material which has been accepted for the award of any other degree or diploma in any tertiary institution, and to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

……

Greg Dutkowski

This thesis may be made available for loan and limited copying in accordance with the Copyright Act 1968.

……

Greg Dutkowski

Abstract

This thesis develops a number of tools and strategies for the adaptation to forest trees of the individual additive genetic model for the prediction of breeding values.

Eucalyptus globulus ssp. *globulus* and central Victorian *E. nitens* are the most important temperate hardwood plantation species in Australia. The geographic patterns of variation in these species were examined using multivariate analysis of open pollinated base population progeny trials. Race classifications were developed from these patterns. New divisions were identified and previously separated provenances were amalgamated. Prediction of breeding values for a variety of traits for *E globulus* showed that the inclusion of races improved the model and increased selection gains by up to 20%.

One problem in the prediction of breeding values in open pollinated base population progeny trials of many genera, including *Eucalyptus*, is that the parents and their offspring do not conform to the assumptions usually made about relatedness in the construction of the additive relationship matrix. An algorithm was developed to modify the additive relationship matrix, and generate its inverse, using simple rules, where parental inbreeding and partial selfing occurs. In simulated data sets, use of the modified relationship matrix lead to unbiased heritability and breeding value estimates. If the correct variance components were used with an incorrect relationship matrix, then the correlation between breeding values was high, but the offspring breeding values were deflated and parental breeding values were inflated.

Breeding value prediction can be further improved by better modelling of environmental variations within trials. The spatial analysis of forest genetic trials using separable autoregressive processes of residuals was adapted from agricultural variety yield trial analysis following the comparison of a number of approaches for five selected forest genetic trials. Augmenting the design model with a spatially auto-correlated component was found to be a good general model which lead to large reductions in design feature effects. The spatial component was found to be relatively small, but with high auto-correlations indicating features spread over relatively large areas. Models without an independent error term were poorer and lead to inflation of estimates of additive variance. The spatial model increased selection gain by up to

6%. Modelling other features identified by the spatial model was not always successful and resulted in only marginal increases in selection gain.

Applying the model to 216 variables from 55 forestry trials resulted in selection gains of more than 10% in around one tenth of cases, although in general the gains were more modest. For growth data, the auto-correlations were generally high, indicating a smooth environmental surface, but they were lower for other traits such a pest and disease damage, indicating more patchiness. Auto-correlations less than zero, indicating competition was dominant, occurred for some large diameter trials, but a bimodal likelihood surface indicated competition was present in more cases. Traits such as stem counts, and form and branching scores, did not respond as often to spatial analysis.

The race classifications, modified relationship matrix, and better environmental modelling developed in the thesis will allow better application of the individual tree additive genetic model to tree breeding programs.

Acknowledgments

I would like to thank my supervisors during my thesis, Arthur Gilmour, Brad Potts, and, early on, Nuno Borralho, for their support in executing this work. I am sure that I have been both the best and worst student for them. Best, as I haven't bothered them too much. Worst, as I haven't bothered them too much as I have been doing lots of other things during the nine years (part-time!) that it has taken me to finish the work that comprises the thesis. So, again, for their support and patience, thank you. I am doubly indebted to Arthur Gilmour for his wonderful software ASReml. Its ease of use has grown over the years, and its power and flexibility remain unsurpassed. It has provided both the inspiration for much of my work, and the means of carrying it out.

While there have been a large number of co-authors in the published work that makes up the thesis, foremost amongst these has been my long time collaborator on the spatial work, João Costa e Silva. Thanks for all the help and stimulating discussion.

As each of the works in the thesis contains its own acknowledgements, I will go no further here than to thank all of the co-authors, technicians, companies and individuals who have designed the experiments and carried out the measurements upon which this thesis work is based. Without their diligent work there would have been no thesis. In quantitative genetics work in forestry it is normally the case that you are analysing someone else's data, saving you the hassle of dusting off your steel-capped boots, and replacing your out-of-date hard hat.

Finally, I would like to thank my family, who have had to put up with me disappearing back to work after dinner and shirking housework for far too long. Mind you, I don't know how much of that will change in the future as there are always more interesting issues to explore.

Other Published Work

While studying part-time I have also been the author or co-author of a number of other works that were not included in this thesis. Below I include a list of all refereed and unrefereed publications that have been produced during the term of this thesis. Those which are part of this thesis are marked with an asterisk.

Refereed Publications

Borralho N.M.G., Dutkowski G.W. (1998) Comparison of rolling front and discrete generation breeding strategies for trees. *Canadian Journal of Forest Research* **28**, 987-993.

Costa e Silva J., Dutkowski G.W., Borralho N.M.G. (2005) Across-site heterogeneity of genetic and environmental variances in the genetic evaluation of *Eucalyptus globulus* trials for height growth. *Annals of Forest Science* **62**, 183-191.

Costa e Silva J., Dutkowski G.W., Gilmour A.R. (2001) Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Canadian Journal of Forest Research* **31**, 1887-1893.

*Dutkowski G.W., Costa e Silva J., Gilmour A.R., Lopez G.A. (2002) Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research* **32**, 2201-2214.

*Dutkowski G.W., Costa e Silva J., Gilmour A.R., Wellendorf H., Aguiar A. (2005) Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Canadian Journal of Forest Research* **submitted**.

*Dutkowski G.W., Potts B.M. (1999) Geographical patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Australian Journal of Botany* **46**, 237-263.

Freeman J.S., Jackson H.D., Steane D.A., McKinnon G.E., Dutkowski G.W., Potts B.M., Vaillancourt R.E. (2001) Chloroplast DNA phylogeography of *Eucalyptus globulus*. *Australian Journal of Botany* **49**, 585-596.

Lopez G.A., Potts B.M., Dutkowski G.W., Apiolaza L.A., Gelid P. (2002) Genetic variation and inter-trait correlations in *Eucalyptus globulus* base population trials in Argentina. *Forest Genetics* **9**, 223-237.

Lopez G.A., Potts B.M., Dutkowski G.W., Traverso J.M.R. (2001) Quantitative genetics of *Eucalyptus globulus*: Affinities of land race and native stand localities. *Silvae Genetica* **50**, 244-252.

Unrefereed Publications

Borralho N.M.G., Dutkowski G.W. (1996) A 'rolling-front' strategy for breeding trees. In proceedings of QFRI - IUFRO Conference 'Tree Improvement for Sustainable Tropical Forestry'. (Eds M.J. Dieters, A.C. Matheson, D.G. Nikles, C.E. Harwood and S.M. Walker). 27 October - 1 November, Caloundra, Queensland. Vol. 2. pp. 317-322. (Queensland Forest Research Institute: Gympie, Queensland)

Costa e Silva J., Dutkowski G.W., Gilmour A.R. (1999) 'Spatial analysis of forest genetic trials.' Internal Report, 16 pp. (Royal Veterinary and Agricultural University: Horsholm, Denmark)

Dutkowski G.W. (2000) '*Eucalyptus globulus* race classification confirmed and extended.' Hot Off the Seed Bed 32, 2 pp. (CRC-SPF: Hobart, Tasmania)

Dutkowski G.W. (2004) Gains from clonal breeding and deployment. In proceedings of IUFRO Conference *'Eucalyptus* in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. p. 163. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Dutkowski G.W. (2004) GSIM - a tree breeding and deployment simulator. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. p. 189. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Dutkowski G.W. (2004) GSOOP - a open-pollinated seed orchard management tool for *Eucalyptus globulus*. In proceedings of IUFRO Conference 'Eucalyptus in a

Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. p. 165. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Dutkowski G.W., Bail I.R., Gore P.L., Powell M.B. (1998) 'Eucalyptus Realised Gains Trials Design.' Technical Report TR98-02. (Southern Tree Breeding Association: Mount Gambier, South Australia)

Dutkowski G.W., Costa e Silva J., Gilmour A.R. (2001) Spatial analysis of forest genetic trials. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. p. 141. (INFOR: Santiago, Chile)

Dutkowski G.W., Gilmour A.R., Borralho N.M.G. (1996) Analysis of growth data using ARMA spatial models. In proceedings of QFRI - IUFRO Conference 'Tree Improvement for Sustainable Tropical Forestry'. (Eds M.J. Dieters, A.C. Matheson, D.G. Nikles, C.E. Harwood and S.M. Walker). 27 October - 1 November, Caloundra, Queensland, Australia. (Queensland Forest Research Institute: Gympie, Queensland)

Dutkowski G.W., Gilmour A.R., Borralho N.M.G. (1997) Application of spatial analysis to tree breeding trials. In proceedings of 'IUFRO Conference on Silviculture and Improvement of Eucalypts'. (Eds A.R. Higa, E. Schaitza and S. Gaiad). 24-29 August, Salvador, Bahia, Brazil. Vol. 1. p. 83. (EMBRAPA: Colombo, Brazil)

*Dutkowski G.W., Gilmour A.R., Borralho N.M.G. (2001) Modification of the additive relationship matrix for open pollinated trials. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. (INFOR: Santiago, Chile)

Dutkowski G.W., Gore P.L., McRae T., Pilbeam D., Apiolaza L. (2000) 'The STBA Cooperative Tree Improvement Strategy for *Eucalyptus globulus*.' Technical Report TR99-03, 30 pp. (Southern Tree Breeding Association/CRCSPF: Mount Gambier, South Australia / Hobart, Tasmania)

*Dutkowski G.W., Potts B.M., Borralho N.M.G. (1997) Revised racial classification of *Eucalyptus globulus* ssp. *globulus* and the importance of including race in analysis

of progeny trials. In proceedings of 'IUFRO Conference on Silviculture and Improvement of Eucalypts'. (Eds A.R. Higa, E. Schaitza and S. Gaiad). 24-29 August, Salvador, Bahia, Brazil. Vol. 1. pp. 322-329. (EMBRAPA: Colombo, Brazil)

Dutkowski G.W., Potts B.M., Williams D.R., Kube P.D., McArthur C. (2001) Geographic genetic variation in Central Victorian *Eucalyptus nitens*. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. p. 39. (INFOR: Santiago, Chile)

Dutkowski G.W., Raymond C.A. (2001) A decision support tool for expensive to measure traits in progeny trials. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. (INFOR: Santiago, Chile)

Dutkowski G.W., Whittock S.P. (2004) 'Gains from Clonal Breeding and Deployment.' 9 pp. (CRC-SPF: Hobart, Tasmania)

Greaves B.L., Dutkowski G.W., McRae T.A. (2004) Breeding objectives for *Eucalyptus globulus* for products other than kraft pulp. In proceedings of IUFRO Conference '*Eucalyptus* in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 175-180. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Greaves B.L., Dutkowski G.W., Raymond C.A., Bamess M., Bulinski J. (2004) Basic density inventory of routine *Eucalyptus globulus* plantations. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 189-190. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Greaves B.L., Hamilton M., Pilbeam D.J., Dutkowski G.W. (2004) Genetic variation in commercial properties of six and fifteen year-old *Eucalyptus globulus*. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 97-102. (RAIZ, Instituto Investigação de Floresta e

Papel: Aveiro, Portugal)

Jordan G.J., Dutkowski G.W., Potts B.M., MacDonald A.C., Tilyard P., Borralho N.M.G. (1998) 'Genetic variation in North Forest Products' *Eucalyptus globulus* ssp. *globulus* base population trials.' Technical Report 8. (Cooperative Research Centre for Sustainable Production Forestry: Hobart, Tasmania, Australia)

Kerr R.J., Dutkowski G.W., Apiolaza L.A., McRae T.A., Tier B. (2002) Developing a genetic evaluation system for forest tree improvement - the making of TREEPLAN®. In proceedings of '7th World Congress on Genetics Applied to Livestock Production'. 19-23 August, Montpellier, France.

Kerr R.J., McRae T.A., Dutkowski G.W., Apiolaza L.A., Tier B. (2001) TREEPLAN - a genetic evaluation system for forest tree improvement. In proceedings of 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. (INFOR: Santiago, Chile)

Kube P.D., Dutkowski G.W. (2002) 'Estimation of *Eucalyptus nitens* breeding values for volume, basic density, pulp yield and branch size at RP252/8 & 9.' Technical Report 28/2002, 21 pp. (Forestry Tasmania: Hobart, Tasmania)

Lopez G.A., Apiolaza L.A., Potts B.M., Dutkowski G.W., Gelid P., Rodriguez Traverso J.M. (2001) Genetic parameters of *Eucalyptus globulus* in Argentina. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. (INFOR: Santiago, Chile)

Lopez G.A., Potts B.M., Dutkowski G.W., Rodriguez Traverso J.M., Gelid P.E. (2001) Quantitative genetic affinities of land races and native stand localities of *Eucalyptus globulus*. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. (INFOR: Santiago, Chile)

McRae T.A., Dutkowski G.W., Pilbeam D.J., Powell M.B., Tier B. (2004) Genetic Evaluation Using the TREEPLAN® System. In proceedings of IUFRO Joint Conference Division 2 'Forest Genetics and Tree Breeding in the Age of Genomics: Progress and Future'. (Eds B. Li and S. McKeand). 1-5 November, Charleston, South

Carolina. (North Carolina State University: Raleigh, North Carolina)

McRae T.A., Pilbeam D.J., Gore P.L., Dutkowski G.W., Apiolaza L.A. (2001) Australian cooperative tree improvement strategy for *Eucalyptus globulus*. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. p. 179. (INFOR: Santiago, Chile)

Pilbeam D.J., Dutkowski G.W. (2004) Simulation to determine optimal wood quality sampling strategies. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 211-212. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Potts B.M., Dutkowski G., Jordan G., Vaillancourt R.E. (1999) Providing a population genetic framework for exploitation of eucalypt genetic resources: The case of *Eucalyptus globulus*. In proceedings of proceedings of '11th Australian Plant Breeding Conference'. (Eds P. Langridge, A. Barr, G. Auricht, G. Collins, A. Granger, D. Handford and J. Paull). Adelaide, South Australia, 19-23 April. Vol. 1 Invited Papers pp. 97-101. (CRC for Molecular Plant Breeding: Adelaide, South Australia)

Potts B.M., Lopez G.A., Volker P.W., Dutkowski G.W., Jordan G.J., Apiolaza L.A., Vaillancourt R.E. (2002) Genetic impacts on the productivity of *Eucalyptus globulus*. In proceedings of 'EucProd 2002: International Conference on Eucalypt Productivity'. 1-5 November 2002, Hobart, Tasmania. p. 28. (CRC for Sustainable Production Forestry: Hobart, Tasmania)

Potts B.M., Milgate A., Joyce K., Mohammed C.L., Vaillancourt R.E., Dutkowski G.W. (2004) Quantitative genetic control of *Mycosphaerella* resistance in *Eucalyptus globulus* and impact on growth. In proceedings of IUFRO Conference 'Eucalyptus in a changing world'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 116-122. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Potts B.M., Vaillancourt R.E.*, et al.* (2004) Exploration of the *Eucalyptus globulus* gene pool. In proceedings of IUFRO Conference 'Eucalyptus in a changing world'.

(Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 46-61. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Whittock S.P., Apiolaza L.A., Dutkowski G.W., Greaves B.L., Potts B.M. (2004) Carbon revenues and economic breeding objectives in *Eucalyptus globulus* pulpwood plantations. In proceedings of IUFRO Conference 'Euclayptus in a changing world'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 146-150. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Table of Contents

Declarations ii

Abstract iii

Table of Contents xiii

List of Figures xvii

List of Tables xx

Chapter 1 Introduction 1

Chapter 2 Geographical patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification 5

2.1 Summary 5

2.2 Introduction 5

2.3 Materials and Methods 8

2.4 Results 15

2.4.1 Spatial Patterns 15

2.4.2 Correlations 20

2.4.3 Race Classification 21

2.5 Discussion 29

2.5.1 Regional Patterns 29

2.5.2 Correlation between Traits 31

2.5.3 Racial Classification 33

2.6 Conclusion 39

2.7 Acknowledgments 39

Chapter 3 Breeding value prediction using the *E. globulus* race classification 40

3.1 Summary 40

3.2 Introduction 40

3.3 Materials And Methods 42

3.4 Results 43

3.5 Discussion 44

3.6 Conclusion 45

3.7 Acknowledgements 45

Chapter 4 Geographic Genetic Variation in, and Race Classification of, Central Victorian *Eucalyptus nitens* 46

4.1 Summary 46

4.2 Introduction 47

4.3 Materials and Methods 48

4.4 Results 53

4.5 Discussion 57

4.6 Conclusion 58

4.7 Acknowledgements 58

Chapter 5 Inversion of the A matrix under parental inbreeding and partial selfing 60

5.1 Summary 60

5.2 Introduction 60

5.3 Materials and Methods 64

5.3.1 Derivation of A^{-1} under partial selfing and parental inbreeding 64

5.3.2 Testing of breeding value predictions 71

5.4 Results 72

5.5 Discussion 76

5.6 Conclusion 78

Chapter 6 Development and comparison of methodology of spatial analysis 79

6.1 Summary 79

6.2 Introduction 80

6.3 Materials and methods 83

6.3.1 Data sets 83

6.3.2 The general statistical model 83

6.3.3 Standard models 85

6.3.4 Variance parameters and model comparison 87

6.3.5 Impact on selection 88

6.3.6 Extended spatial models 88

6.4 Results 93

6.4.1 Standard models 93

6.4.2 Extended spatial models 96

6.4.3 Variance parameters and accuracy of breeding value predictions 99

6.4.4 Breeding value correlations and relative genetic gains 103

6.5 Discussion 103

6.6 Conclusion 107

6.7 Acknowledgements 107

Chapter 7 Application of spatial analysis methodology to many data sets 108

7.1 Summary 108

7.2 Introduction 109

7.3 Materials and methods 111

7.3.1 Data sets 111

7.3.2 Model fitting 117

7.4 Results 122

7.5 Discussion 135

7.6 Conclusion 141

7.7 Acknowledgements 141

Chapter 8 Concluding Discussion 143

References 154

List of Figures

Figure 2-1 Location of seed collection localities with respect to the species distribution. 10

Figure 2-2 Geographic variation of traits. Locality means of synthetic, leaf morphology and other traits. 18

Figure 2-3 Locality scores on the major discriminant functions. 24

Figure 2-4 Geographic variation of discriminant scores..... 25

Figure 2-5 Average linkage clustering of localities. 28

Figure 2-6 The revised racial classification. 32

Figure 4-1 Distribution of *Eucalyptus nitens* and *E. denticulata* in Australia. 47

Figure 4-2 Species distribution, sub-provenances, and localities of central Victorian *Eucalyptus nitens*. 49

Figure 4-3 Discriminant function scores for localities on the first three axes. 56

Figure 4-4 Average linkage clustering of the significant discriminant axis scores and the proposed races. 57

Figure 5-1 The effect of parental inbreeding (F) and selfing rate on the parent-offspring relationship. 66

Figure 5-2 The effect of parental inbreeding (F) and selfing rate on the offspring-offspring relationship. 66

Figure 5-3 The effect of parental inbreeding (F) and selfing rate on the offspring inbreeding (F)..... 67

Figure 5-4 Additive variance estimates at different rates of selfing. 73

Figure 5-5 Error variance estimates at different rates of selfing. 73

Figure 5-6 Heritability estimates at different rates of selfing. 74

Figure 5-7 Correlation (r) between true and predicted parental breeding values for different rates of selfing. 74

Figure 5-8 Correlation (r) between true and predicted offspring breeding values for different rates of selfing. 75

Figure 5-9 Slope (b) of the relationship between true and predicted parental breeding values for different rates of selfing. 75

Figure 5-10 Slope (b) of the relationship between true and predicted offspring breeding values for different rates of selfing. 76

Figure 5-11 Gain relative to the KNOWN model for different rates of selfing. 77

Figure 6-1 Simulated AR1⊗AR1 process for a 40 x 40 grid. 90

Figure 6-2 Simulated AR1⊗AR1 process with independent error for a 40 x 40 grid. 91

Figure 6-3 Data, best base model surface and variogram. 95

Figure 6-4 Extended model surfaces and variograms. 98

Figure 7-1 Height and diameter (0.3 or 1.3m) growth for each trial. 116

Figure 7-2 Changes in log likelihood (ΔLogL) from adding within replicate design features and spatial components to the simplified design model for more complex trial designs. 124

Figure 7-3 Spatial and independent error variance in spatial model by trait type. ... 125

Figure 7-4 Distribution of auto-correlations by trial shape and trait type. 126

Figure 7-5 Spatial (Sp) and extended spatial (Sp+) models for variables with highly asymmetric auto-correlations. 127

Figure 7-6 Effect of spatial analysis on variance due to design factors by trait type.

.....130

Figure 7-7 Genetic variance (additive or clonal) for design and spatial models by trait type.....131

Figure 7-8 Change in additive genetic value accuracy for spatial models by trait type. 132

Figure 7-9 Correlation between breeding values from the design and spatial models for parents and offspring by trait type.....133

Figure 7-10 Gain from selection for spatial models by trait type. 134

Figure 8-1 *Eucalyptus globulus* locality variation on the major discriminant functions from Dutkowski (2000a). 143

Figure 8-2 UPGMA clustering of *E. globulus* native stand and land race localities from Lopez *et al.* (2001). 144

Figure 8-3 Revised race and sub-race boundaries. 147

Figure 8-4 Comparison of molecular and quantitative affinities between races. 149

List of Tables

Table 2-1 Localities sampled in the 1987 and 1988 CSIRO collections of *Eucalyptus globulus* that are represented in this analysis. 9

Table 2-2 Description of traits measured. 12

Table 2-3 Locality ANOVAs and locality mean spatial clustering. 16

Table 2-4 Principal components analysis of locality means and their spatial clustering. 17

Table 2-5 Correlations of trait locality means. 17

Table 2-6 Correlation between locality trait means and climatic variables. 20

Table 2-7 Discrimination between localities and their spatial clustering. 21

Table 2-8 Allocation of localities to race and sub-race groups. 34

Table 2-9 Summary of attributes of major races from our data and other published information. 36

Table 3-1 The effect of including race into the prediction of breeding values. 43

Table 4-1 The geographic origin of localities sampled and families used in the trials. 50

Table 4-2 Distribution of families across the trials and trial design information. 51

Table 4-3 Traits measured and their means. 52

Table 4-4 Probability that there are no locality mean differences. 54

Table 4-5 Correlations between locality means of synthetic traits. 55

Table 5-1 Rules for creation of the inverse NRM under parental inbreeding and selfing and comparison with Henderson's (1976) rules. 70

Table 6-1 Trial and trait information. 84

Table 6-2 Akaike Information Criterion relative to the base model. 94

Table 6-3 Variance parameter estimates and breeding value accuracies for selected standard models and extended spatial models. 100

Table 6-4 Correlation of breeding values and gains in selection between models... 103

Table 7-1 Summary of trials used in the analysis. 112

Table 7-2 Data analysed (other than growth) grouped by type. 114

Table 7-3 Distribution of changes in log likelihood with spatial model by trait type. 123

Table 7-4 Distribution of changes in log-likelihood with respect to the ratio of the sum of design feature variances to the error variance in the design model and the number of observations. 124

Table 8-1 Localities sampled in revisions of the *E. globulus* race classification. 145

Chapter 1 Introduction

Breeding values, or additive genetic values, represent the average additive effects of the genes that an individual receives from its parents (Falconer and MacKay 1996). It is the part of genetic variation that can be used for population improvement through selection. In progeny trial analysis the challenge is to appropriately separate environmental from genetic effects so that selections can be made for future breeding and deployment. In animal breeding, the method of Henderson called best linear unbiased prediction (BLUP) has been increasingly used for the prediction of breeding values and other genetic effects, simultaneously with the estimation of other fixed and random effects in the framework of the Mixed Model Equations (MME) (Henderson 1984; Mrode 1998). The properties of BLUP are neatly incorporated into its name:

- Best – maximises the correlation between true and predicted random effects,
- Linear – predictors are a linear function of the data,
- Unbiased – the expectations of the fixed and random effects are unbiased, and
- Prediction – it involves the prediction of random effects.

The use of the Additive (or Numerator) Relationship Matrix, which takes the co-ancestry between trees into account, in BLUP further enables it to use all additive relationships to allow estimation of breeding values of all individuals in the pedigree and to take selection into account in both the estimation of variances and prediction of breeding value, if the selection data is included (Henderson 1975).

BLUP differs from BLP (Best Linear Prediction) in that the fixed effects are estimated, rather than being assumed to be known. Both can be used for the prediction of random effects, although BLUP has often come to mean prediction of breeding values through the use of the Numerator Relationship Matrix. White and Hodge (1989) argue that in many forestry situations fixed effects can be adequately estimated so the computational complexity of BLUP can be avoided. Certainly in many situations the problems that led to the development of BLUP for animal breeding – highly unbalanced data making it difficult to estimate sub-class means,

genetic trend due to multiple generations of data, and more data on better animals due to culling – are not present in base generation tree breeding programs with large balanced trials. For animal breeding, prediction of breeding values requires the use of large data sets because of the lack of structure and experiments in the breeding populations. This has not been the case for trees.

Increasingly, however, breeding value prediction in forestry trials is moving closer to the animal breeding model. Programs are moving into advanced generations where simple trial means are no longer unbiased estimates of site effects, and relatedness and selection need to be taken into account in variance component estimation and breeding value prediction. Trial designs are now more complex than the simple randomised complete block designs of the past with cyclic and computer-generated designs (Nguyen and Williams 1993) using incomplete blocks within replicates and models using recovery of inter-block information (Williams and Matheson 1994). The computational limitations of the past are being removed with the general increase in processing speed and the development of software such as ASReml (Gilmour *et al.* 1999) for variance component estimation and solving of the mixed model equations. These changes are gradually seeing an increase in the use of the mixed model equations and the numerator relationship matrix in the prediction of breeding values (Jarvis *et al.* 1995; Araújo *et al.* 1997; Fernandez *et al.* 1998; Soria *et al.* 1998; Wei and Borralho 2000; Apiolaza and Garrick 2001).

In their comparison of BLP and BLUP, White and Hodge (1989) urged that the assumptions used in the application of BLUP to animal breeding be examined for their appropriateness when the method is used in tree breeding. Specifically they raise the issues of heterogenous variances, estimation of population effects, dealing with inbreeding and coancenstry, the effect of selection and computational feasibility. Similarly, Borralho (1995) argued that the problems due to uncertain pedigree in open-pollinated material, spatial auto-correlation, and heterogenous variances amongst classes of fixed effects all required attention. This thesis attempts to examine certain aspects of the use of BLUP in tree breeding to ensure that the models that are used are the most appropriate.

Appropriate definition of genetic groups is vital to the best estimation of variance

components and prediction of breeding values. Chapters 2 to 4 examine the geographic patterns of variation in two plantation eucalypt species that have worldwide importance, *Eucalyptus globulus* and *E. nitens*. Based on multivariate variation, race boundaries are proposed, and the effect of using races on model fit, variance component estimates and breeding value predictions is examined.

Often in forest trees we are dealing with populations that have been sampled from the wild which violate some of the assumptions usually made in the construction of the Numerator Relationship Matrix for animals. The parents are often inbred and the seed from them is subject to partial selfing which can lead to problems in variance component estimation (Squillace 1974; Askew and El-Kassaby 1994; Borralho 1994). In Chapter 5, the Numerator Relationship Matrix is modified to cope with parental inbreeding and partial selfing, and rules for calculation of its inverse are derived. The effect of using or ignoring these assumptions on variance component estimation and breeding value prediction is then examined.

Spatial analysis as a means of better separating environmental from genetic or other treatment effects in analysis of designed trials has been growing in agriculture for some years (Braysher *et al.* 2001; Singh *et al.* 2003; Yang *et al.* 2004). Chapter 6 compares different spatial models for their utility in a number of selected tree breeding trials and recommends a general approach based on model improvement, the ease of model fitting and gains from selection. Chapter 7 then applies this method to a large range of data sets in order to examine how widespread its utility is, and appreciate the likely gains in selection that can be made. The final chapter attempts to put the thesis in the context of other recent work.

Most of the work presented in this thesis has been previously published in journal articles and conference proceedings – only the expanded section on the derivation of the inverse of the Numerator Relationship Matrix in Chapter 5 has been added. As each chapter is a published work, they have been altered little for the thesis, except to homogenise format and numbering, and to have a single list of references. As the works are largely self contained, the introduction and discussion for the thesis is suitably brief, and there is a small degree of unavoidable repetition in the introductions to each work. As some of the work was published some years ago, the

final chapter also attempts to put the work in a current context. I have also elected to retain the use of "we" in references to the authors of each work as each of these were the results of collaborations with both my academic supervisors and others.

Chapter 2 Geographical patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification*

2.1 Summary

The geographic patterns of genetic variation in a wide variety of quantitative traits were studied in *Eucalyptus globulus* ssp. *globulus* and its intergrades, leading to a revised racial classification. The analysis was based on 35 traits assessed across five field trials in northern Tasmania from approximately 500 open-pollinated families, encompassing 49 collection localities in native stands. There were significant differences between the collection localities for most traits. While growth and survival exhibited weak spatial structuring, there were clear regional patterns in bark thickness, wood basic density, flowering precocity and some aspects of juvenile leaf morphology. There were a number of significant correlations between trait locality means, but few simple correlations between the regional patterns observed and climatic variables. Multivariate analyses indicated that the localities could be effectively amalgamated into larger, geographically concordant races. A hierarchy of five major groupings of 13 races and 20 subraces is proposed to account for most of the quantitative genetic variation while allowing for outliers and intermediate populations. Some areas of the distribution may need further sampling to more accurately elucidate their racial affinities, especially those with traits of high economic importance.

2.2 Introduction

Spatially structured genetic variation within a species can arise where genetic drift or adaptive differentiation are not overcome by the homogenising effects of dispersal

---

* Published as Dutkowski, G.W. and Potts, B.M. (1999). Geographical patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Australian Journal of Botany* **46**, 237-263.

and gene flow (Endler 1977; Roughgarden 1979). Such is likely to be the case for widespread plant species with no special mechanisms for dispersal, or for which barriers to dispersal exist. Natural selection may also act on a number of different traits, in different places and at different times, resulting in extant patterns of spatial genetic variation reflecting both current and historic patterns of differentiation. Spatial patterns of genetic variation may thus be complex, especially where changing environments have caused changes in selection pressures, dispersal barriers and species distribution.

Regardless of the cause of genetic differentiation, the existence of spatial structure in genetic variation on a relatively broad scale allows the delineation of geographic races, within which individuals are more related to each other than to individuals in other groups (Endler 1977). Such genetic variation is often multi-dimensional and continuous in nature, and the exact delineation of races may be difficult if clear geographical and genetic disjunctions are not present. Nevertheless, racial groups are of interest to breeders because they standardise the classification of seed sources, enable links between different seed collections, provide a means of summarising complex patterns of genetic variation, and may improve the prediction of breeding values through the use of a genetic groups model (Quaas 1981). An earlier version of the classification derived in this paper has been shown to improve the accuracy of breeding value prediction and thus improve the gains from selection by up to 20% (Dutkowski *et al.* 1997).

Eucalyptus globulus Labill. ssp. *globulus*, a native of south-eastern Australia, is likely to exhibit spatial genetic variation. It has a widespread distribution, occurs in a number of different environments, and is likely to have had a complex history of environmental change, migration, barriers to dispersal and intergradation with related subspecies (Jordan *et al.* 1993; Potts and Jordan 1994a). Although its pollen dispersal mechanisms are undocumented, its relatively heavy seed means that seed dispersal is likely to be limited (Cremer 1977). On the basis of phenotypic variation in floral morphology and climatic information, Kirkpatrick (1975) amalgamated *E. globulus* Labill. and its related species into a single species with subspecies *globulus*, *bicostata*, *maidenii* and *pseudoglobulus* (hereafter referred to by their subspecific names) with

extensive zones of intergradation. Jordan *et al.* (1993) extended this work, recognising a number of core and intergrade zones between the subspecies.

Eucalyptus globulus ssp. *globulus* is one of the most widespread plantation pulpwood species in the world, with over 1,700,000 ha planted, and over 70,000 ha planted in Australia by 1996 (Tibbits *et al.* 1997). There are active breeding programs in Australia (Butcher 1990; Jarvis *et al.* 1995), Chile (Prado and Alvear 1993), Portugal (Borralho and Cotterill 1994), Spain (Vega Alonso *et al.* 1994) and China (Zang *et al.* 1995). In many of these programs the subspecies is still in the early stages of domestication, so the investigation of the patterns of genetic variation in native stands is still important. Genetic differences have been revealed when seed collected from native stands has been grown under more or less uniform conditions in field trials. Most studies of ssp. *globulus* and its intergrades have been based on trials established from two major seed collections (Orme 1977; Gardiner and Crawford 1987, 1988). Spatially structured genetic variation has been reported for survival (Almeida 1993; Prado and Alvear 1993), growth (Volker and Orme 1988; Almeida 1993; Prado and Alvear 1993; Spencer and Williams 1993; Potts and Jordan 1994b; Vega Alonso *et al.* 1994; Kube *et al.* 1995; Zang *et al.* 1995), taper (Guimaraes *et al.* 1995), pilodyn penetration (MacDonald *et al.* 1998), stem form (Zang *et al.* 1995), fungal resistance (Carnegie *et al.* 1994; Zang *et al.* 1995), insect feeding (Farrow *et al.* 1994), stem borer resistance (Soria and Borralho 1998), frost resistance (Tibbits *et al.* 1991; Almeida *et al.* 1995; Zang *et al.* 1995), drought resistance (Dutkowski 1995; Toro *et al.* 1998), juvenile leaf morphology (Potts and Jordan 1994a), persistence of juvenile foliage (Spencer and Williams 1993; Potts and Jordan 1994a), seedling abnormalities (Potts and Jordan 1994b), and variation in RAPD markers (Nesbitt *et al.* 1995).

Jordan *et al.* (1994) derived a racial classification of the subspecies and its intergrades based on the offspring from open-pollinated seed collected from trees throughout the natural range grown to age 4 years in five trials in Tasmania. Although there was clearly spatially structured variation, much of the variation was continuous and the racial groups were not distinct. Discontinuities in the geographic distribution had to be used to define races. Some localities represented by only a few families also proved difficult to classify, despite their geographic proximity to other localities. A

classification based on RAPD variation (Nesbitt *et al.* 1995) showed poor geographical clustering of localities, although a major latitudinal cline was detected.

Data from the same field trials used by Jordan *et al.* (1994) are now available for numerous traits other than growth, such as survival, flowering precocity, wood basic density, and juvenile leaf morphology, many of which are of economic importance. A geographical race classification should be most effective when it is based on the genetic variation of many traits which exhibit spatial variation. Incorporation of economic traits ensures the relevance of the races to current breeding objectives, while incorporation of a variety of non-economic traits helps ensure that the racial groups are robust. This paper summarises the spatial patterns of genetic variation of these traits and examines their relationships with each other and with environmental variables. The patterns of variation are used to revise the racial classification of this taxon.

2.3 Materials and Methods

The families studied were growing in field trials established from a range-wide collection of open pollinated seed from native stand parent trees of *globulus* and its intergrade populations made in 1987 and 1988 by the Australian Tree Seed Centre of the CSIRO (Gardiner and Crawford 1987, 1988). Progeny trials of this material have been planted in Australia, Chile, Portugal, Spain and China (MacDonald *et al.* 1995; Zang *et al.* 1995) and have formed the basis of a number of breeding programs. A varying number of easily accessible, representative trees were collected from areas within the range of the subspecies. Potts and Jordan (1994b) grouped the collection areas into 46 localities, which were an arbitrarily defined area of approximately 10 km diameter. These localities, and three extra localities with few families used in this study, are detailed in Table 2-1 and shown in Figure 2-1.

Table 2-1 Localities sampled in the 1987 and 1988 CSIRO collections of *Eucalyptus globulus* that are represented in this analysis.

The codes follow Jordan *et al.* (1994) up to code 46; three extra localities (47-49) have been included. The number of families in the trials and their representation in the two classification data sets (LAMGEX and ALL) analysed are shown. Locality 14 was not included in the classification analysis as it was not sampled for all traits.

Code	Locality	Number of Families			
		Range in Trials	Any Trial	LAMGEX Data	ALL Data
1	South West Lavers Hill	3-6	6	5	2
2	Otway State Forest	28-43	44	39	23
3	Cannan Spur	18-21	21	21	18
4	Parker Spur	37-56	59	51	35
5	Cape Patton	9-18	21	15	7
6	Jamieson Creek	7-7	7	7	7
7	Lorne	16-17	17	16	16
8	Jeeralang North	48-51	51	49	45
9	Jeeralang	2-3	3	3	2
10	Madalya Road	6-8	9	8	5
11	Bowden Road	5-5	5	5	5
12	Port Franklin	2-4	5	2	1
13	Hedley	7-12	13	8	6
14	Wilsons Promontory Lighthouse	12-16	16	-	-
15	North Flinders Island	12-13	14	12	12
16	Central North Flinders Island	8-13	13	9	7
17	Central Flinders Island	13-23	23	20	13
47	Central East Flinders Island	0-1	1	1	0
18	South Flinders Island	10-12	13	10	10
19	North Cape Barren Island	10-10	10	10	10
20	West Cape Barren Island	27-34	34	32	27
21	Clarke Island	6-6	6	6	6
22	St. Helens	6-11	11	10	6
23	Pepper Hill	5-10	10	8	4
24	Royal George	6-9	10	9	5
25	German Town	5-5	5	5	4
26	Mayfield	2-5	6	4	2
27	Taranna	3-5	5	5	3
28	Triabunna	3-9	10	6	3
29	North Maria Island	5-7	7	6	5
30	Mt. Dromedary	4-4	4	4	3
31	Ellendale	4-4	5	4	4
32	Jericho	8-10	10	10	5
33	Collinsvale	4-5	5	4	4
34	Hobart South	9-10	10	9	9
35	Moogara	20-26	26	24	19
36	Blue Gum Hill	3-4	4	4	2
37	South Geeveston	7-7	7	7	5
48	North Geeveston	1-3	3	3	1
49	Strathblane	1-1	1	1	1
38	Dover	3-5	6	4	2
39	South Bruny Island	2-7	7	6	2
40	Recherche Bay	2-4	5	4	2
41	Port Davey	3-6	6	3	2
42	Macquarie Harbour	4-8	8	7	4
43	Little Henty River	11-11	11	11	11
44	Badgers Creek	8-10	10	9	8
45	South King Island	10-10	10	10	10
46	Central King Island	18-22	23	21	17
Total			616	527	400

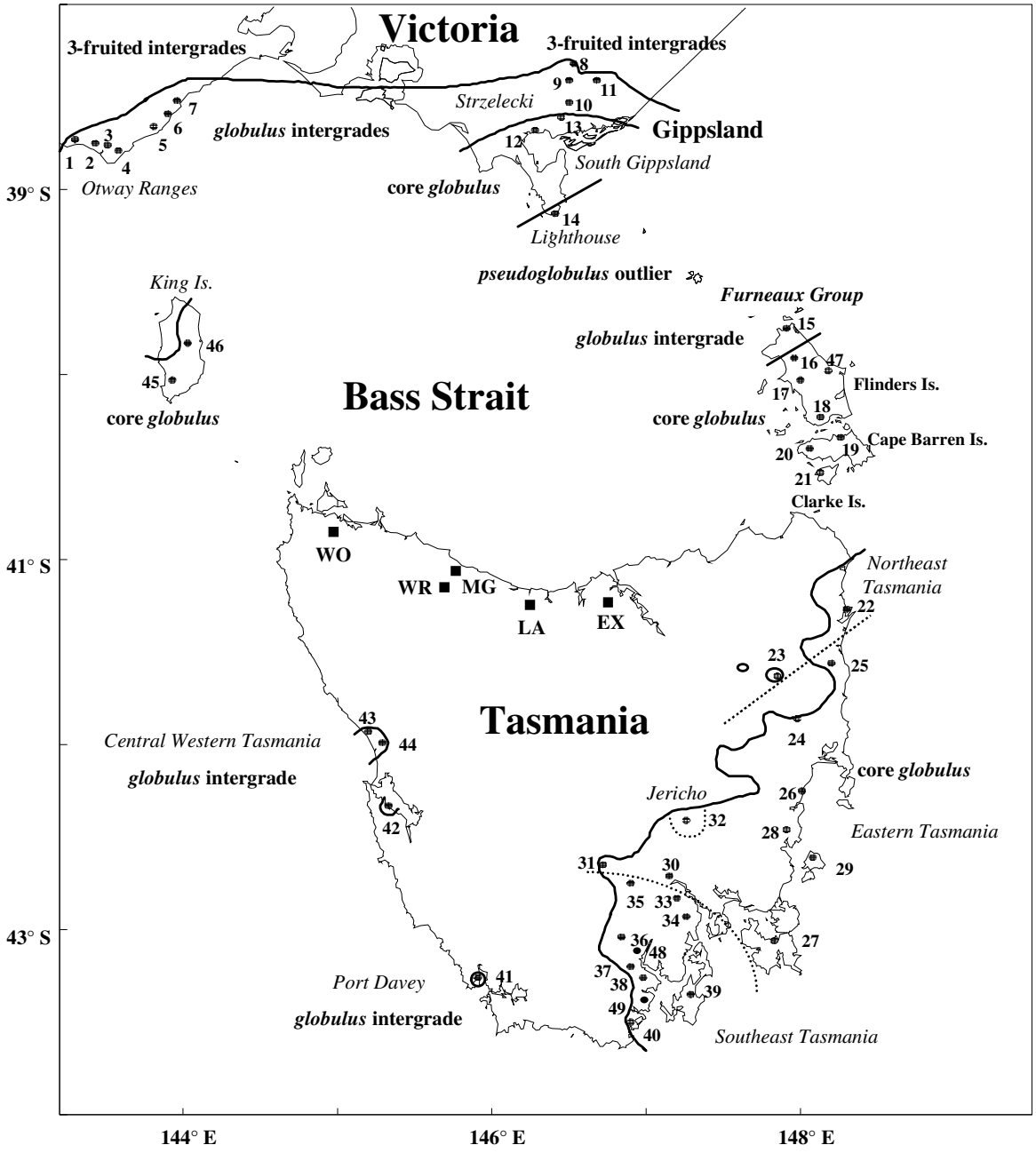


Figure 2-1 Location of seed collection localities with respect to the species distribution.

Symbols are (●) seed collection localities (codes follow Table 2-1); (—) distribution and subspecies boundaries of Jordan *et al.* (1993); (▪▪▪▪▪) race boundaries of Jordan *et al.* (1994); (■) trial sites with codes used in the text.

The five progeny trials used in this analysis were established by North Forest Products in the northwest of Tasmania in 1989; Exeter (EX), Latrobe (LA), Massy Greene (MG), West Ridgley (WR) and Woolnorth (WO). The trials encompassed a wide range of environments suitable for plantations of the subspecies in north-western Tasmania, but all fell outside its natural distribution (Figure 2-1). Each trial was established with row plots of two trees in five complete replicates with between 21 and 28 incomplete blocks in a resolvable incomplete block design. The number of families at any one trial varied between 450 and 596 (see Jordan *et al.* (1994) for further details).

Direct and indirect measurements were made of growth, survival, bark thickness, flowering precocity, wood basic density, leaf damage by sawfly (*Perga affinus*) larvae, and juvenile leaf morphology and persistence (Table 2-2). Growth was measured as height and diameter. Bark thickness was calculated as a proportion of diameter. Pilodyn penetration was used as an index of wood basic density (Greaves *et al.* 1996). Not all traits were measured at all sites, nor were all trees measured for each variable. Only the latest, or most complete, measurement of a variable was used, except for bark thickness, where an earlier, more complete, measurement at Massy Greene was also included. The measurements of the 12 traits on between one and five sites yielded 35 variables.

Data from runts, multi-stemmed or damaged trees, and measurements deemed to be outliers were rejected. The data for some variables were subject to a transformation (see Table 2-2) to ensure normality and homogeneity of variance. Family least square means were calculated with the univariate linear model:

[2-1] $Y = \mu + Rep + Inblk + Fam + \varepsilon$

where **Y** is the observation of the variable for the tree, μ is the mean of the variable, **Rep** is the replicate as a fixed effect, **Inblk** is the incomplete block as a random effect, **Fam** is the open-pollinated family as a fixed effect, and ε is the random error.

Table 2-2 Description of traits measured.

The traits measured, the code used in subsequent tables, trait description, the trials in which the data was collected, the age at which it was collected, the tree sampling strategy, and the transformation used in the analysis.

Trait	Code	Trials	Description	Age (years)	Sample	Trans-formation
Height	HT	All	Height to tallest growing tip	4.0	All trees	Square
Diameter	DBH	All	Diameter at Breast Height (1.3m).	4.0	All trees	
Pilodyn penetration	PILO	All	Average of 2 shots in a small bark window on the western side of the tree at 1.3m	5.5	First tree in each plot from 2 replicates	
Bark thickness	BRK4	MG	Double bark thickness, average of 4 measurements around the stem at 1.3m using a bark thickness gauge, as a proportion of DBH	4.0	1 tree per plot in 4 replicates	Ln
Bark thickness	BRK5	All	Double bark thickness from 1 measurement in the window used for pilodyn measurement, as a proportion of DBH	5.5	First tree in each plot from 2 replicates	Ln
Survival	SURV	All	Survival as a binary trait	4.0	All trees	
Flowering precocity	PREC	EX MG WR	Flowering precocity: presence (1) or absence (0) of flower buds or capsules at a young age	4.0	All trees	
Leaf length	LL	MG	Leaf length of the average juvenile leaf sampled at 1.3m	1.5	All trees	Ln
Leaf width	LW	MG	Leaf width, as a proportion of LL	1.5	All trees	Ln
Length to the widest point	LWP	MG	Length to widest point on leaf, as a proportion of LL	1.5	All trees	Square root
Basal lobing	BASE	MG	Length of basal lobe on leaf, as a proportion of LL	1.5	All trees	Square root
Height to phase change	HTPC	MG	Height on main stem to first petiolate leaf, as a proportion of tree height	1.5	All trees from 4 replicates	Square root
Sawfly damage	SAW	MG	Defoliation by Sawfly larvae (*Perga affinus* subsp. *insularis*, PERGIDAE) scored on a three point scale	5.5	All trees	

The family least square means were calculated using VCE version 3.2 (Groeneveld 1996) and PEST version 3.0 (Groeneveld 1990), SAS (SAS Institute 1990), or ASReml (Gilmour *et al.* 1997b). The incomplete block term was not included in the model for precocity as insufficient trees were sampled (or flowered) in each block to reliably estimate this effect.

The family least square means were used in one-way ANOVAs to test for differences between locality means. The geographic pattern of variation and the correlations of the traits with each other, and with climatic variables, were examined using locality means. The tendency of the locality means to display a spatial structure was assessed by a measure of spatial autocorrelation. In general, the approach used to classify localities into races followed Jordan *et al.* (1994) involving multivariate discrimination and clustering techniques. Wilsons Promontory Lighthouse (14; hereafter termed Lighthouse) was generally not included in the analyses because of its extremely slow growth.

In order to summarise the geographic variation, locality means (including Lighthouse (14) for comparison) were plotted against their geographic position for each trait. To summarise variation in traits measured by more than one variable, synthetic traits were created from locality means for groups of similar variables: growth (using height and diameter), bark thickness (using BRK4 and BRK5), pilodyn penetration, survival and flowering. These synthetic traits were the first axis from principal components analyses based on the correlations between the means of the 33 localities with at least three families for each variable used. Scores on the first axis were then calculated for these 33 localities and an extra 11 localities with fewer families. The principal components analysis was undertaken using Genstat 5.32 (Payne *et al.* 1988).

Pairwise Pearson's correlations of the localities means were calculated between the traits, and between the traits and selected climatic variables. The climatic variables were selected as they summarised the mean and variation in annual rainfall and temperature and as they showed little correlation with each other. The climatic variables used were mean annual temperature (tann), the difference between the coldest monthly mean minimum and the warmest monthly mean maximum (tspan), the mean annual rainfall (rann), and the coefficient of variation of mean monthly

rainfall (rcvar) (see Jordan *et al.* (1993) for further details). Correlations significantly different from zero were detected using both a single comparison and a multiple comparison test (using Bonferroni's inequality) (Snedecor and Cochrane 1980). Only the 44 localities with trait means based on more than three families were used.

Discriminant function analysis (DFA) (Sokal and Rohlf 1981) was used to discriminate between localities. DFA finds linear combinations of family means that maximally differentiate localities in multivariate space. Locality discriminant scores for all localities were calculated from the discriminant coefficients based on a subset of the data with only those localities with more than three families for all the variables. The discriminant scores were graphed for the first three discriminant functions and mapped in their geographic position for the first four functions.

For the DFA, only families measured for all variables could be used. The main data set used (denoted ALL) comprised 35 variables measured on the five sites for 372 families from 33 localities. In order to sample a greater number of families and localities a second data set was formed with just those families common to the sites with the greatest number of families: LA, MG, and EX. This data set (denoted LAMGEX) comprised 24 variables measured on the three sites for 513 families from 42 localities.

To group the localities into races, the locality discriminant scores were used to hierarchically classify the localities. The primary clustering was done using average linkage clustering (Sneath and Sokal 1973) of all discriminant functions for only those localities with more than three families; however, the clustering was also carried out with different sets of localities (all localities, and only localities with more than three families), using different clustering techniques (average linkage, single-linkage and Ward's method), and with different numbers of discriminant functions (all, or only the significant discriminant functions). The grouping of localities into races was done primarily on the clustering, but was supplemented by the maps of the locality discriminant scores and other information about specific localities. The discriminant function analysis and clustering were undertaken using SAS (SAS Institute 1990).

The tendency of the locality trait means and discriminant scores to have similar values for localities that are closer together (to be spatially clustered or autocorrelated)

was determined in order to verify that there was a spatial structure to the variation. A variogram describes the spatial autocorrelation in a data set by calculating the variance of differences between pairs of values at different distances apart (Isaaks and Srivastava 1989). If there is a tendency for spatial autocorrelation, then the variogram will increase with distance. The significance of the increase with distance was determined by an F-test of an isotropic bounded linear model fitted to the observed variogram in 10 km distance classes using Genstat 5.32 (Payne *et al.* 1988).

2.4 Results

2.4.1 Spatial Patterns

There were significant locality differences for all variables, except for survival (SURV) at EX and LA (Table 2-3). There was significant spatial autocorrelation of locality means in all groups of variables, but not for all variables within a group (Table 2-3). Pilodyn penetration (PILO) at WR showed the highest spatial autocorrelation, while diameter (DBH) at MG showed the lowest. The PILO and bark thickness (BRK) variables showed the strongest spatial autocorrelation, while survival (SURV) showed the lowest. Leaf length (LL) and leaf width (LW) were strongly autocorrelated, and length to widest point (LWP) and basal lobing (BASE) were not.

The first principal component axis explained a large proportion of the variation of the locality means for most groups of variables (Table 2-4). The first principal components for flowering precocity and bark thickness each explained over 80% of the variation, indicating that these traits were very stable across trial sites. Survival was the least stable trait across sites, with less than half of the variation explained by the first principal component axis. Spatial autocorrelation of the synthetic traits was again strongest for pilodyn and bark thickness, and was not significant for growth and survival.

Table 2-3 Locality ANOVAs and locality mean spatial clustering.

For locality ANOVA, the *F* ratio and probability of no difference between localities from univariate ANOVA of family means, and for spatial clustering, the *F* ratio and probability that a bounded linear model explains none of the observed variogram.

Trait	Site	ANOVA		Spatial Clustering	
		F-ratio	*F*-Prob.	*F*-ratio	*F*-Prob.
HT	EX	3.05	< 0.001	0.72	0.491
	LA	3.57	< 0.001	3.22	0.047
	MG	5.91	< 0.001	1.26	0.291
	WO	6.67	< 0.001	3.07	0.054
	WR	4.28	< 0.001	1.68	0.195
DBH	EX	2.75	< 0.001	2.20	0.120
	LA	2.93	< 0.001	3.12	0.052
	MG	4.50	< 0.001	0.01	0.990
	WO	6.08	< 0.001	3.63	0.033
	WR	3.04	< 0.001	1.18	0.315
BRK4	MG	3.73	< 0.001	3.34	0.042
BRK5	EX	6.00	< 0.001	5.30	0.008
	LA	8.17	< 0.001	3.94	0.025
	MG	5.26	< 0.001	2.95	0.061
	WO	3.62	< 0.001	4.03	0.023
	WR	3.68	< 0.001	3.54	0.036
PILO	EX	4.89	< 0.001	4.30	0.018
	LA	4.91	< 0.001	4.77	0.012
	MG	4.71	< 0.001	2.11	0.131
	WO	3.41	< 0.001	3.29	0.044
	WR	3.97	< 0.001	7.57	0.001
SURV	EX	1.00	0.484	0.25	0.779
	LA	1.16	0.221	2.97	0.059
	MG	2.95	< 0.001	1.09	0.343
	WO	3.50	< 0.001	1.82	0.171
	WR	5.40	< 0.001	2.40	0.099
PREC	LA	5.59	< 0.001	3.31	0.043
	MG	7.05	< 0.001	2.76	0.072
	WR	3.53	< 0.001	2.92	0.062
LL	MG	8.43	< 0.001	4.56	0.015
LW	MG	3.52	< 0.001	4.95	0.010
LWP	MG	2.33	< 0.001	1.28	0.286
BASE	MG	2.07	< 0.001	0.40	0.672
HTPC	MG	6.20	< 0.001	2.58	0.084
SAW	MG	3.02	< 0.001	3.19	0.049

Table 2-4 Principal components analysis of locality means and their spatial clustering.

The variables used to create synthetic traits and the proportion of variation explained by the first principal component (PCA-1), and for the spatial clustering, the *F* ratio and probability that a bounded linear model explains none of the observed variogram.

Synthetic Trait	Variables Used	PCA-1	Spatial Clustering	
		(%)	*F* ratio	*F* Prob.
Growth	HT, DBH	54.9	1.40	0.254
Bark	BRK4,BRK5	82.0	6.95	0.002
Pilodyn	PILO	71.7	7.70	0.001
Survival	SURV	45.4	0.94	0.397
Precocity	PREC	88.4	3.37	0.041

Table 2-5 Correlations of trait locality means.

Correlations of locality means between the synthetic traits and the variables, leaf morphology (LL, LW, LWP & BASE), height to phase change (HTPC), and sawfly damage (SAW). Individual significant correlations are marked as $^{+}$ $p< 0.05$, $^{++}$ $p < 0.01$, $^{+++}$ $p < 0.001$, or significant correlations detected using Bonferroni's inequality within the whole suite of correlations are marked as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Synthetic Traits					Variables				
	Growth	Bark	Pilo-dyn	Survi-val	Preco-city	LL	LW	LWP	BASE	HTPC
Bark	-0.08									
Pilodyn	0.08	-0.42^{++}								
Survival	0.56^{**}	-0.11	0.12							
Precocity	0.15	0.16	-0.13	0.41^{++}						
LL	-0.04	0.14	0.12	-0.32^{+}	-0.50^{*}					
LW	0.02	-0.26	-0.17	0.11	0.20	-0.57^{**}				
LWP	-0.47^{+++}	-0.10	-0.12	0.35^{+}	-0.21	-0.19	0.44^{++}			
BASE	0.39^{++}	-0.33^{+}	0.19	0.51^{*}	0.30^{+}	-0.33^{+}	0.38^{++}	-0.42^{++}		
HTPC	0.26	0.09	-0.02	-0.11	-0.33^{+}	0.65^{***}	-0.19	-0.25	0.01	
SAW	-0.04	0.54^{**}	0.09	0.05	0.34^{+}	-0.02	-0.37^{+}	-0.42^{++}	0.01	-0.11

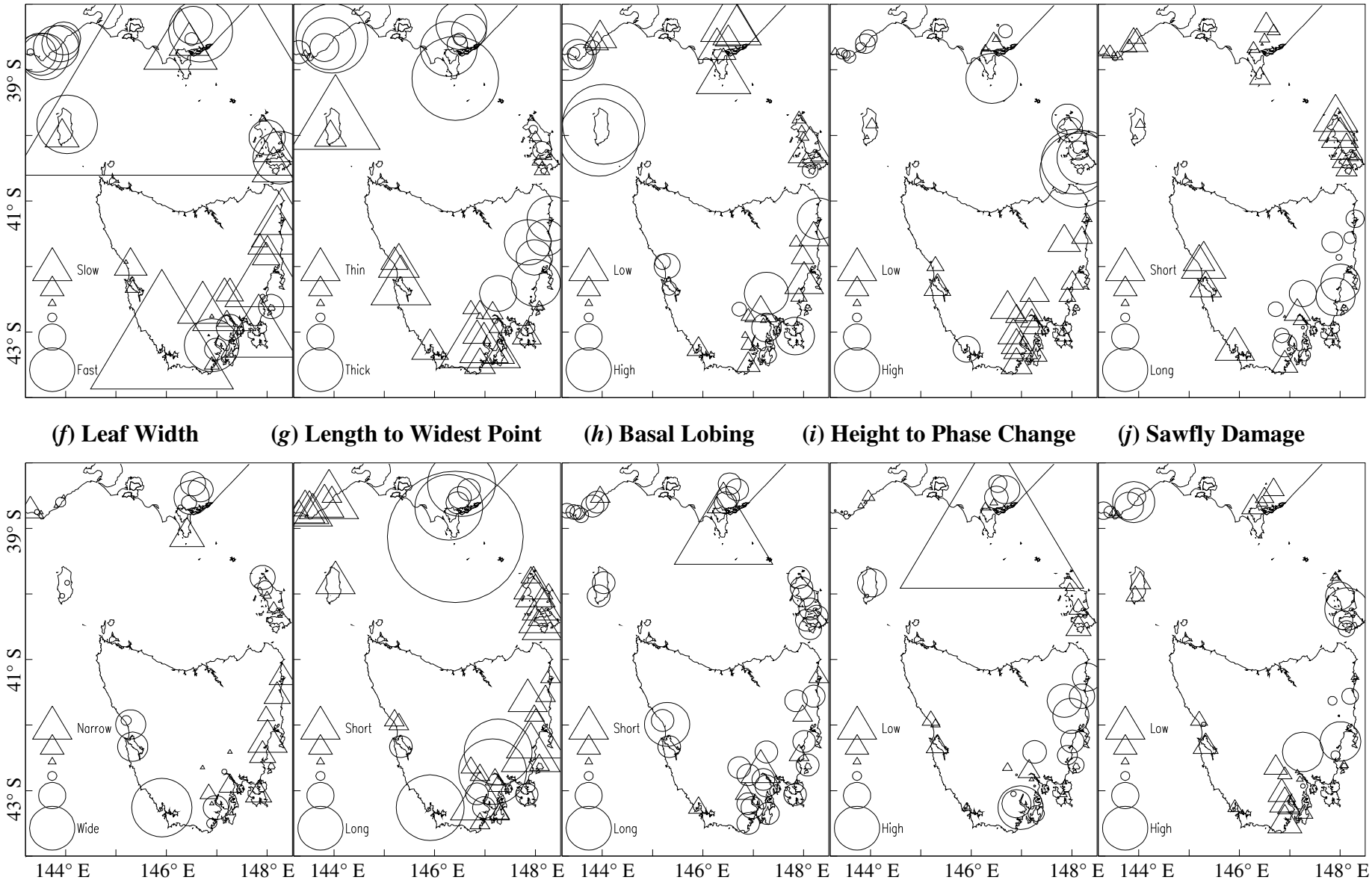


Figure 2-2 Geographic variation of traits. Locality means of synthetic, leaf morphology and other traits.

The geographic patterns of variation of the synthetic traits (except survival), juvenile leaf morphology, sawfly damage and height to phase change are summarised in Figure 2-2. For growth (Figure 2-2a), while the spatial autocorrelation was not significant, the Otway and Strzelecki Ranges, and southern Tasmania were regions of above average growth. Lighthouse (14) and Port Davey (41) were slow growing, but localities from eastern Tasmania and south Gippsland also showed below average growth, while the Bass Strait islands were variable. Bark thickness exhibited clear regional differences (Figure 2-2b). Gippsland, the eastern Otway Ranges, and north-eastern Tasmania had thick bark. There was a steep cline in bark thickness in the Otway Ranges from west to east, and from north to south in the more or less continuous distribution along the east coast of Tasmania. The Furneaux Group and western Otways were intermediate, while King Island had thin bark. The lowest pilodyn penetration was found in Gippsland, with below-average penetration in the eastern Otway Ranges, the Furneaux Group, and some parts of eastern Tasmania (Figure 2-2c). The highest penetration was found at King Island, with high penetration also in the western Otway Ranges, the central west coast of Tasmania, and isolated localities in eastern Tasmania. The Furneaux Group, especially the southern islands, and Lighthouse (14), were by far the most precocious flowering regions (Figure 2-2d). Port Davey and the Otway Ranges were above average, but there was little early flowering elsewhere. Long juvenile leaves were found in eastern Tasmania (Figure 2-2e), short leaves in western Tasmania and the Furneaux Group, with intermediate leaf lengths elsewhere. The widest leaves were from western Tasmania and Gippsland (except Lighthouse (14)) (Figure 2-2f), the narrowest leaves from eastern Tasmania and Lighthouse (14), and localities elsewhere were intermediate. The length to the widest point was high in Gippsland and isolated localities in Tasmania (Figure 2-2g). Basal lobes were notably short at Lighthouse (14) (Figure 2-2h), but no clear pattern was evident otherwise. The least juvenile foliage by far was retained by Lighthouse (14) (Figure 2-2i), while the west coast of Tasmania, south-eastern Tasmania, and the Furneaux Group were below average. More juvenile foliage than average was retained by localities in the Strzelecki Ranges, and the northern and southern parts of the Tasmanian east coast distribution. Sawfly damage was high for the eastern Otways (Figure 2-2j), the Furneaux Group and the central east coast of Tasmania, low for Gippsland, King Island, and western and south-

eastern Tasmanian, and intermediate elsewhere. Survival (not shown in Figure 2-2 because of its low percentage of variation explained by the first principal component axis and its lack of spatial autocorrelation) showed no strong patterns, with the localities with the maximum and minimum survival occurring in close proximity in Gippsland.

2.4.2 Correlations

The locality means (Table 2-5) show the strongest correlations (significant at 5% by the Bonferroni test) between leaf length (LL) and height to phase change (HTPC) (0.65), leaf length (LL) and leaf width (LW) (-0.57), growth and survival (0.56), bark thickness and sawfly damage (SAW) (0.54), and survival and basal lobing (BASE) (0.51). For these major correlations there was little evidence that geographic groups of high leverage inflated the correlation estimates. A number of correlations between traits were found that were only significant ($P < 0.05$) by a univariate test.

Table 2-6 Correlation between locality trait means and climatic variables.

Correlations of locality means of synthetic traits and variables, leaf morphology (LL, LW, LWP & BASE), height to phase change (HTPC), and sawfly attack (SAW), with climatic variables, mean annual temperature (tann), the difference between the coldest months mean minimum temperature and the warmest months mean maximum temperature (tspan), the mean annual rainfall (rann), and the coefficient of variation of mean monthly rainfall (rcvar). Individual significant correlations are marked as $^{+}$ $p < 0.05$, $^{++}$ $p < 0.01$, $^{+++}$ $p < 0.001$, or significant correlations detected using Bonferroni's inequality within the whole suite of correlations are marked as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	tann	tspan	rann	rcvar
Growth	0.23	0.28	-0.15	0.35^{+}
Bark	0.28	0.56**	-0.32^{+}	-0.23
Pilodyn	-0.03	-0.49*	0.12	0.45^{++}
Survival	0.00	0.03	-0.03	0.32^{+}
Precocity	0.61***	0.08	-0.07	0.29
LL	-0.42^{++}	0.07	-0.30^{+}	-0.27
LW	0.21	-0.10	0.52**	0.19
LWP	-0.12	0.01	0.29	-0.16
BASE	0.23	-0.24	0.16	0.39^{++}
HTPC	-0.23	0.14	-0.13	-0.19
SAW	0.28	0.20	-0.31^{+}	0.03

The strongest correlations with climatic variables were between flowering precocity and annual temperature (tann) (0.61), bark thickness and the temperature span (tspan) (0.56), leaf width (LW) and rainfall (rann) (0.52), and pilodyn penetration and tspan (–0.49) (Table 2-6). Of these, all but the correlation between annual tspan and bark thickness seem to be primarily caused by regional groups of data points of high leverage. For the correlation of tann with precocity it was the relatively warm and precocious Furneaux Group, for pilodyn penetration and tspan it was the low tspan and high penetration of King Island, and for leaf width (LW) and rann it was the high rainfall and wide leaves of western Tasmania.

Table 2-7 Discrimination between localities and their spatial clustering.

For each of the two data sets, the proportion of locality differences explained by the discriminant functions, and for the spatial clustering, the *F* ratio and probability that a bounded linear model explains none of the observed variogram of the discriminant scores.

Discriminant Function	ALL Data			LAMGEX Data		
	Variation Explained (%)	Spatial Clustering		Variation Explained (%)	Spatial Clustering	
		F ratio	*F* Prob.		*F* ratio	*F* Prob.
1	26.5	5.44	0.007	27.1	3.24	0.046
2	18.0	6.26	0.003	16.9	6.17	0.004
3	12.3	2.46	0.095	12.9	2.66	0.079
4	9.3	4.49	0.015	9.0	3.57	0.035

2.4.3 Race Classification

The first three discriminant functions for all the variables explained just over half of the locality variation for both data sets (Table 2-7). Significant ($P < 0.05$) variogram models indicated spatial autocorrelation of the localities on all but the third function (Table 2-7). This was reflected in the strong regional patterns evident in the graphs of the first three discriminant functions (Figure 2-3) and the maps of the discriminant scores (Figure 2-4) even for localities with few families. The regional groups resulting from the two data sets (Table 2-3) were very similar, as were the variables contributing to their separation, although the direction of the vectors for the variables and the positions of the localities differed by a rotation of about 45° between the two data sets. Vectors for groups of variables used in the principal components analysis tended to lie in the same direction, except for SURV. The low between-locality

F-ratio for SURV (Table 2-2), however, indicated that it contributed little to discriminating between localities. Eastern Tasmania and King Island were clearly separated from other populations on a combination of LL, HTPC, PREC and height at Massy Greene (MG-HT). The differences within these groups were generally the result of differences in BRK and PILO.

Localities from the western Otway Ranges and the Furneaux Group were the closest groups on the first two discriminant functions in the ALL data set, being separated by the third function in both data sets and the second function in the LAMGEX data set. They were close to the southern Gippsland localities of Madalya Road (10) and Hedley (13), which were separated from the geographically close Strzelecki Ranges localities. The eastern Otways were well separated from the western Otways on the first discriminant function, although the geographically intermediate Cape Patton locality (5) was also intermediate on this function and switched its closest affinity between the east and west Otways groups in the two data sets. The eastern Otways were closest to the Strzelecki Ranges. King Island was clearly separated from the Furneaux Group on the first function. The Tasmanian localities formed three major groups on the second function: northern and southern groups on the east coast, and a western group. Recherche Bay (40) was intermediate between the southern east coast group and the western group in both data sets, as was Port Davey (41) in the ALL data. The southern east-coast group was further divided into a southern and a south-eastern group on the first and third functions.

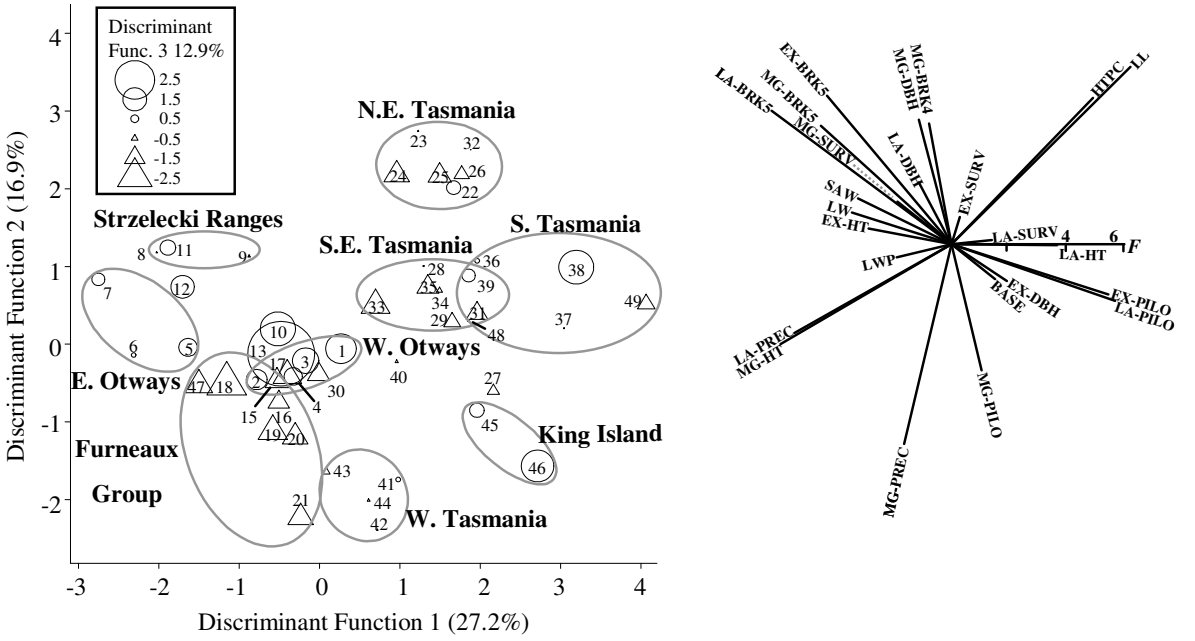
There were few localities that did not conform well to the patterns observed, and these were generally represented by few families. Mt. Dromedary (30) was an outlier from the south-eastern Tasmania group in the LAMGEX data, lying closer to the western Otways. Port Franklin (12), a locality represented by two or less families, was an outlier to the Hedley (13) and Madalya Road (10) groups, lying closer to the Strzelecki Ranges in the LAMGEX data, and with no clear affinity in the ALL data. Taranna (27) was closer to King Island than to the south-eastern group, especially in the LAMGEX data.

The hierarchical clustering showed a strong spatial grouping of localities for the average linkage clustering of localities with more than three families (Figure 2-5) as well as for the other clustering procedures examined (not shown). The patterns were

similar to those shown in the graphs and maps of the discriminant functions (Figure 2-3 and Figure 2-4); however, differences did occur, presumably because the first three discriminant functions explained only half of the variation. For both data sets the clustering showed five major groups: Victoria, the Furneaux Group, King Island, western Tasmania and eastern Tasmania. The western Otways, however, changed its closest affinity from the eastern Otways in the LAMGEX data, to the Furneaux Group in the ALL data, and King Island showed a close affinity to western Tasmania in the LAMGEX data. Within these major groups, the western and eastern Otways formed separate groups, although Cape Patton changed its affinity between these groups in the two data sets, as it had done on the graphs of the discriminant functions (Figure 2-3). The southern islands of the Furneaux Group (Clarke and Cape Barren) formed a subgroup in the LAMGEX data, coincident with the cline on the second discriminant function shown in Figure 2-3a and Figure 2-4a. Within eastern Tasmania there was a stable south-eastern group evident, and two north-eastern groups (German Town (25) and Royal George (24), and Jericho (32) and Pepper Hill (23)) in both data sets. Two southern localities (South Geeveston (37) and Dover (38)) also clustered together in the LAMGEX data. The affinities of Taranna (27) with King Island, and Recherche Bay (40) with western Tasmania were shown in the LAMGEX data.

There were a number of localities, however, that did not cluster according to any spatial pattern. Mt. Dromedary (30), Hedley (13), and Mayfield (26) in the LAMGEX data set, and Hedley (13) and St. Helens (22) in the ALL data set, were unrelated to any group. South West Lavers Hill (1) was an outlier to the mainland–Furneaux cluster in the LAMGEX data, and Clarke Island (21) and South Flinders Island (18) were outliers to the western Tasmania-western Otways-Furneaux Group cluster in the LL data, as was Madalya Road (10) to the mainland-western Tasmania cluster. South Bruny Island (39) and Taranna (27) clustered with spatially unrelated groups in the LAMGEX data.

(*a*) LAMGEX Data Set



(*b*) ALL Data Set

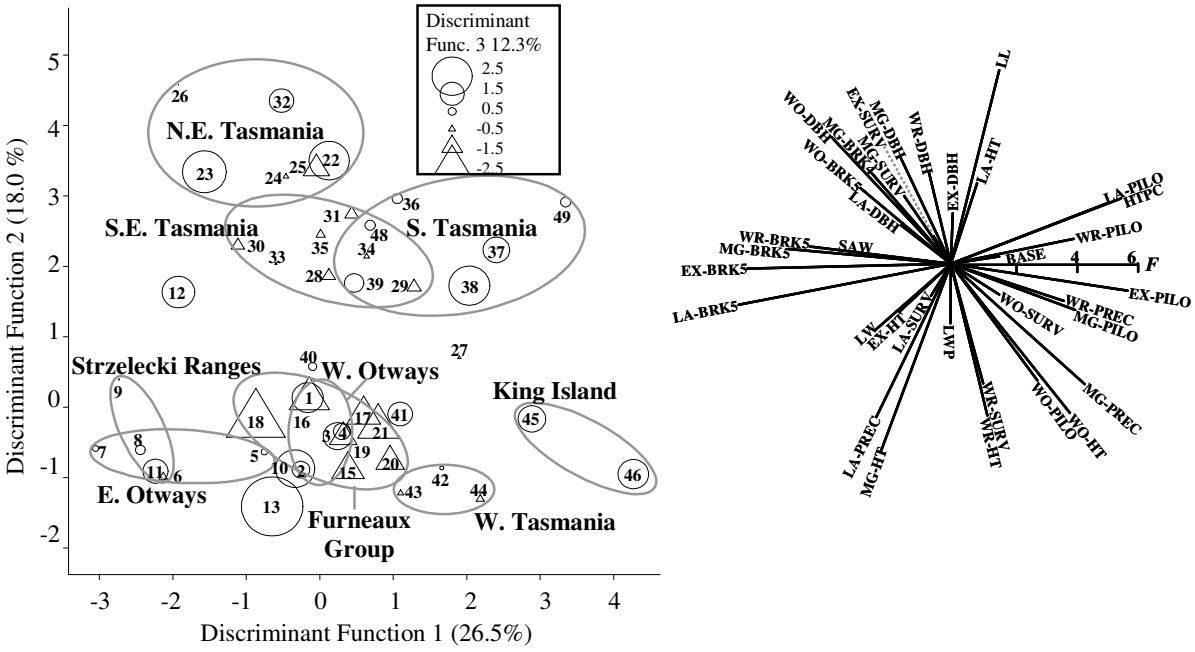


Figure 2-3 Locality scores on the major discriminant functions.

Locality scores for the first three discriminant functions for the (a) LAMGEX and (b) ALL data sets. The locality code from Table 2-1 is shown for each data point. The third discriminant function is shown by the size of the symbols for each locality; circle for positive scores and triangles for negative scores. The vectors on the right show the direction of the trait discriminant coefficients, with the length proportional to the between-locality F-ratio. The trait codes are the same as in Table 2-2.

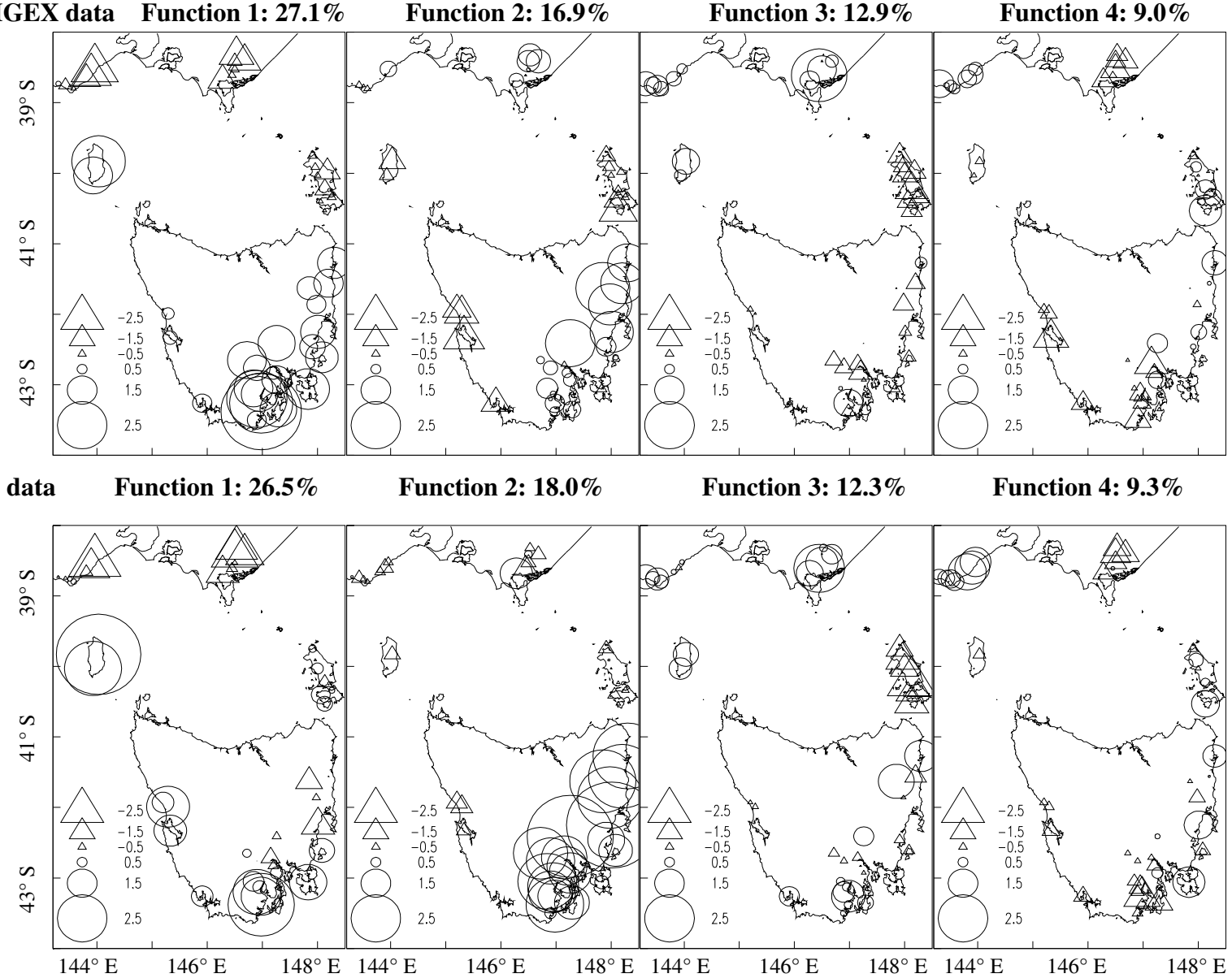


Figure 2-4 Geographic variation of discriminant scores.

Locality discriminant scores for discriminant functions 1 to 4, showing the proportion of variation accounted for by each function.

The strong spatial structure evident in the discriminant scores and the hierarchical clustering indicated that a geographically based racial classification was possible. As the variation was both hierarchical and clinal in nature, and there was no clear break in the clustering hierarchy, localities were assigned to a hierarchy of races and sub-races. The races were designed to summarise major regional groups of localities and strongly differentiated individual localities. Sub-races catered for localities that formed less differentiated outliers from the races, or that were intermediate between the races, as well as to account for spatial or physiographic discontinuities. Localities with few families were allocated to races or subraces primarily on the basis of the graphs and maps of the discriminant scores, as they often clustered with unrelated groups.

The division of the localities into races was predominantly from the clusters fused at a Mahalanobis' distance of less than 3.5 for both data sets (Figure 2-5). The races proposed for the mainland were the Eastern and the Western Otways, Strzelecki Ranges, Southern Gippsland, and Wilsons Promontory Lighthouse (Figure 2-6 and Table 2-8). The Otway Ranges were split into two races as there were major differences between the two groups on the first discriminant function, as well as for economically important traits such as pilodyn penetration and drought susceptibility (Dutkowski 1995), and they are geographically separate along a rainfall gradient. The Strzelecki Ranges form a clear group, genetically and physiographically separated from Southern Gippsland which is on a coastal plain. Port Franklin (12) was included in the Southern Gippsland race even though it was somewhat different from Hedley (13). These two localities are geographically close on the coastal plain, but Port Franklin is represented by only a small number of families. Cape Patton (5) was treated as a sub-race within the Eastern Otways race because of its geographic and genetic intermediacy between the Western and Eastern Otways races. Madalya Road (10) was treated as a sub-race within the Strzelecki Ranges race as, although it had strong affinities with Southern Gippsland, it clustered with the Strzelecki Ranges race, and was likely to be genetically intermediate between the two. It is also physiographically between the two, being located in the foothills of the Strzelecki Ranges, close to the coastal plain of southern Gippsland. The Furneaux Group was clearly a separate and distinct race with close affinities to mainland races. A

Southern Furneaux sub-race was recognised, largely because of its differentiation in the LAMGEX data, which is probably a result of its more precocious flowering (Figure 2-2d) and smaller juvenile leaves (Figure 2-2e). While Wilsons Promontory Lighthouse (14) was not included in the clustering, it was sufficiently different from its neighbouring localities for growth, flowering precocity, leaf morphology, retention of juvenile foliage, and RAPD markers (Nesbitt *et al.* 1995) to be treated as a separate race.

King Island was treated as a separate race, with its closest affinities to Southern and Western Tasmania. Western Tasmania was also considered a distinct race, with affinities to the mainland races and Southern Tasmania. Port Davey (41) was genetically close to the rest of the west coast on the major discriminant functions; however, it was somewhat genetically different, and seemed to form part of a cline from the upper west coast localities around to the southern localities (it is also close to Recherche Bay (40) on RAPD data (Nesbitt *et al.* 1995)), and as it is geographically isolated, it was established as a separate sub-race. On the Tasmanian east coast the grouping of localities into regions was more difficult because of the more continuous nature of the genetic variation and geographical distribution. There were, however, clear groups with sufficient disjunction to separate them. The area was predominantly divided into North-eastern, South-eastern and Southern Tasmania races.
Mt. Dromedary (30) and Recherche Bay (40), although represented by few families, were both considered sufficiently different from their neighbouring localities to merit allocation to separate races at this stage. These Mt. Dromedary (30) families have been shown to have a high proportion of seedlings with green and sub-glaucous foliage with leaf shape indicative of advanced generation hybridisation with *E. ovata*, or similar species (Potts and Jordan 1994b), a characteristic common to juvenile foliage on trees in the area (Brad Potts, *pers. comm.*). Recherche Bay (40) was treated as a separate race because, although it is geographically closer to southern Tasmanian localities, it was intermediate between the Southern and Western Tasmanian races on the discriminant scores, clustered closer to the Western Tasmanian race (Figure 2-5a), and was close to Port Davey (41) in analyses of RAPD variation (Nesbitt *et al.* 1995).

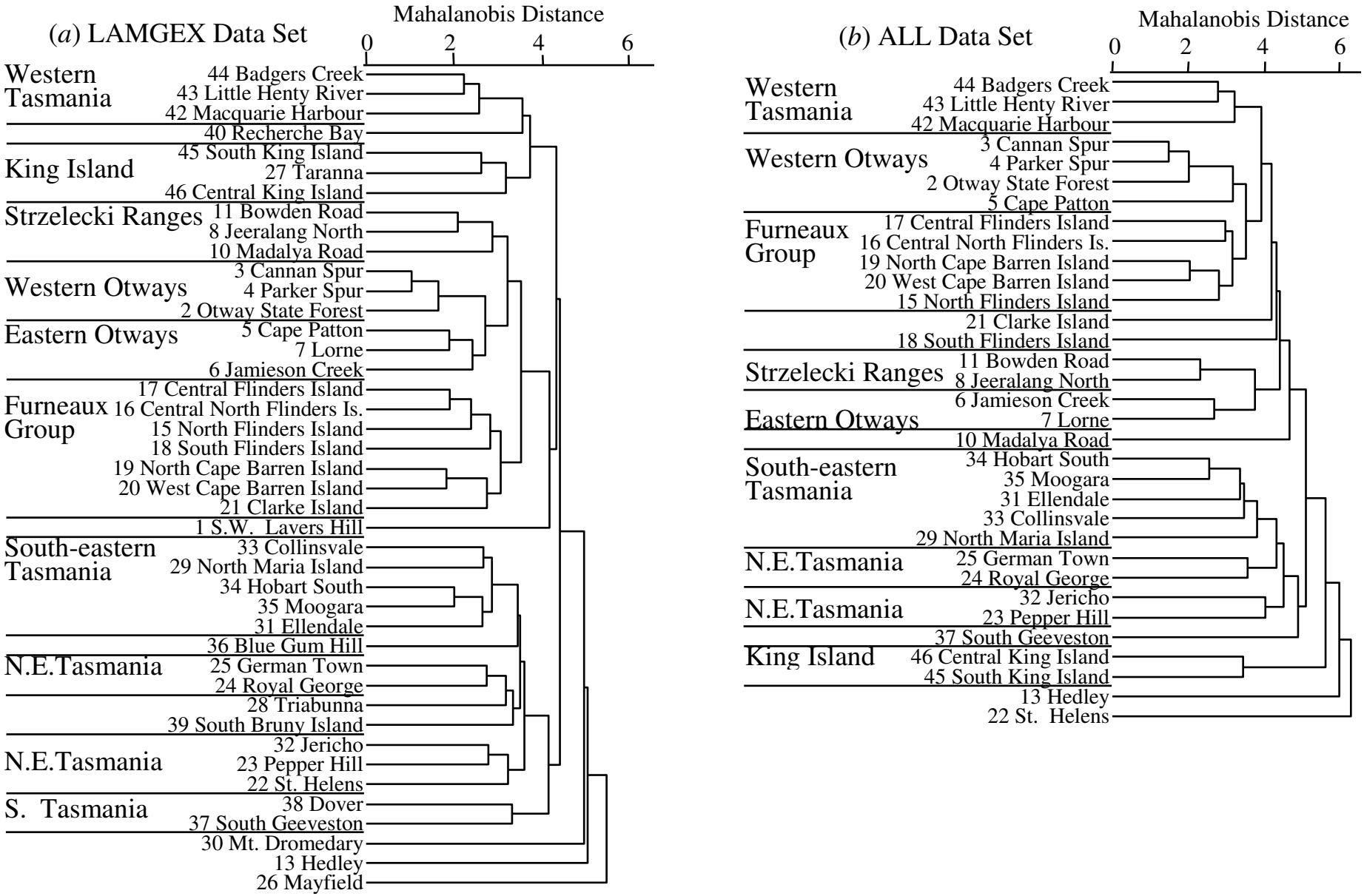


Figure 2-5 Average linkage clustering of localities.

Clustering of the (a) LAMGEX and (b) ALL data sets based on Mahalanobis distances between localities from discriminant function analysis for localities with more than three families.

Although the North-eastern Tasmania race forms a fairly stable and distinct group on the major discriminant functions, with a major disjunction on the second discriminant function separating it from more southerly localities (Figure 2-4b), the clustering (Figure 2-5) shows the race to be quite diverse, with two major subgroups. Jericho (32) and Pepper Hill (23) were separated as an Inland North-eastern Tasmania sub-race because of their stable clustering. St. Helens (22) was also separated out as a sub-race as it was an outlier in the ALL clustering. Mayfield (26) was kept within the core North-eastern sub-race, even though it was an outlier in the clustering shown, as it has a similar discriminant value to that group on function 2 in the LAMGEX data, marking the edge of the major north–south disjunction on that function in eastern Tasmania. The South-eastern Tasmania race was separated from the Southern race by its differences on the first and third discriminant functions and the consistent clustering of the core south-eastern group. Triabunna (28) was included in the South-eastern race, to which it is geographically close, and with which it has its closest affinity on the major discriminant functions, even though it clustered with some North-eastern Tasmania race localities in the LAMGEX data. The Southern Tasmania race formed a group on the major discriminant functions and is a distinct geographic area in the wet southern forests of Tasmania, but formed no major hierarchical cluster. The Tasman Peninsula, although represented by few families, was a separate sub-race within the southern race. It may be closer in affinity to King Island, but its next closest affinity was to the Southern Tasmania race localities, to which it was geographically close but somewhat isolated by a large bay.

2.5 Discussion

2.5.1 Regional Patterns

Eucalyptus globulus ssp. *globulus* and its intergrades clearly show a regional pattern of quantitative genetic variation which make them amenable to racial classification. Such a pattern of variation can arise from historic patterns of migration, adaptation and genetic drift. This analysis cannot effectively discriminate between these mechanisms. While there is a geographic structure to the genetic variation, there is still substantial within-locality genetic variation for growth (Potts and Jordan 1994b),

leaf morphology (Potts and Jordan 1994a) and pilodyn penetration (MacDonald *et al.* 1998) which has not been explored here. Despite regional variation in most traits, few simple relationships with broad-scale climatic factors were found. The relationship may be more complex (Potts and Jordan 1994a), or involve environmental variables other than the macro-climatic ones examined. Adaptation may be occurring at scales smaller than the locality level that we used (approximately 10 km). Clines over 2 km have been observed in coastal areas (Chalmers 1992). Such clines would not, however, account for the regional patterns observed. The widespread and disjunct distribution of the species could have led to differential adaptation to environmental factors within different regions. There is an east-west cline in bark thickness and drought tolerance (Dutkowski 1995) in the Otway Ranges coincident with a decline in rainfall. This could indicate an adaptation of thicker bark to a higher fire frequency, and of tolerance to water stress. Nevertheless, such a relationship between bark thickness and rainfall cannot be seen over the whole of the range of the subspecies. The north-south cline on the east coast of Tasmania also suggests adaptation to some environmental change coincident with latitude; however, no clear association with macro-climatic variables is evident.

The complex migratory history of the subspecies suggested by Jordan *et al.* (1993) could well mean that the current patterns are a melange of current and past patterns of adaptation. Jordan *et al.* (1993) suggested, on the basis of capsule morphology and a more likely eastern land bridge across Bass Strait, that the core *globulus* subspecies type may have originated in Tasmania and migrated to Victoria, primarily via the Furneaux Group, where it underwent secondary intergradation with three fruited types. While the RAPD data (Nesbitt *et al.* 1995) would support such a hypothesis, placing King Island, the Furneaux Group and northern Tasmania between Victoria and southern Tasmania, our data indicate a major genetic discontinuity between northern Tasmania and the geographically nearby Furneaux Group (Figure 2-3 and Figure 2-4). This may, however, indicate adaptation to environments subject to exposure to strong westerly winds in the isolated populations of the Furneaux Group, King Island and western Tasmania.

2.5.2 Correlation between Traits

Correlation between population means can arise through common (pleiotropic) or linked gene action, through co-adaptation, or indeed by chance alone. In this study, it was not possible to differentiate between these mechanisms. The few correlations with environmental variables detected here do not provide much evidence for co-adaptation. The correlation between leaf length and height to phase change at the between-population level was also found by Potts and Jordan (1994a), but its absence at the within-population genetic level led them to hypothesise that parallel variation in these traits may result from parallel adaptation to the same environmental gradients. The positive association between survival and growth suggests that growth rate may be related to overall fitness at these mesic trial sites. Chambers *et al.* (1996) found a moderate within-race genetic correlation between survival and growth, which would support such a hypothesis. However, on a broader geographic scale there may be sites where genes for fast growth are unrelated to, or have a negative effect on, long-term plant fitness. For example, on more marginal sites, other traits (such as bark thickness to aid in recovery from fire, or drought tolerance) may contribute more to fitness. However, it is possible that there is a stable component to the relationship between growth rate and survival arising from differences between localities in inbreeding levels (Potts and Jordan 1994b). Indeed, the impact of factors influencing inbreeding, such as population density and size (Borralho and Potts 1996), may mask environmental factors and explain the poor spatial structuring observed for growth and survival traits. The correlation between leaf length and leaf width may simply indicate an allometric relationship whereby absolute leaf width is relatively constant, or increases more slowly than leaf length, hence causing a negative correlation between leaf length and the relative leaf width we used. The low correlation observed between bark thickness and pilodyn penetration could be due to the joint origin of wood and bark in the cambium, and hence may be pleiotropic. However, the correlation between bark thickness and sawfly damage cannot be so easily explained (sawflies attack the leaves, not the bark, although larvae do move about the stem) nor can that between bark thickness and basal lobing of the leaves. Examination of within-population genetic correlations may shed light on the nature of the correlations; those due to pleiotropy, as opposed to genetic disequilibrium, should

also be present at the within-population level.

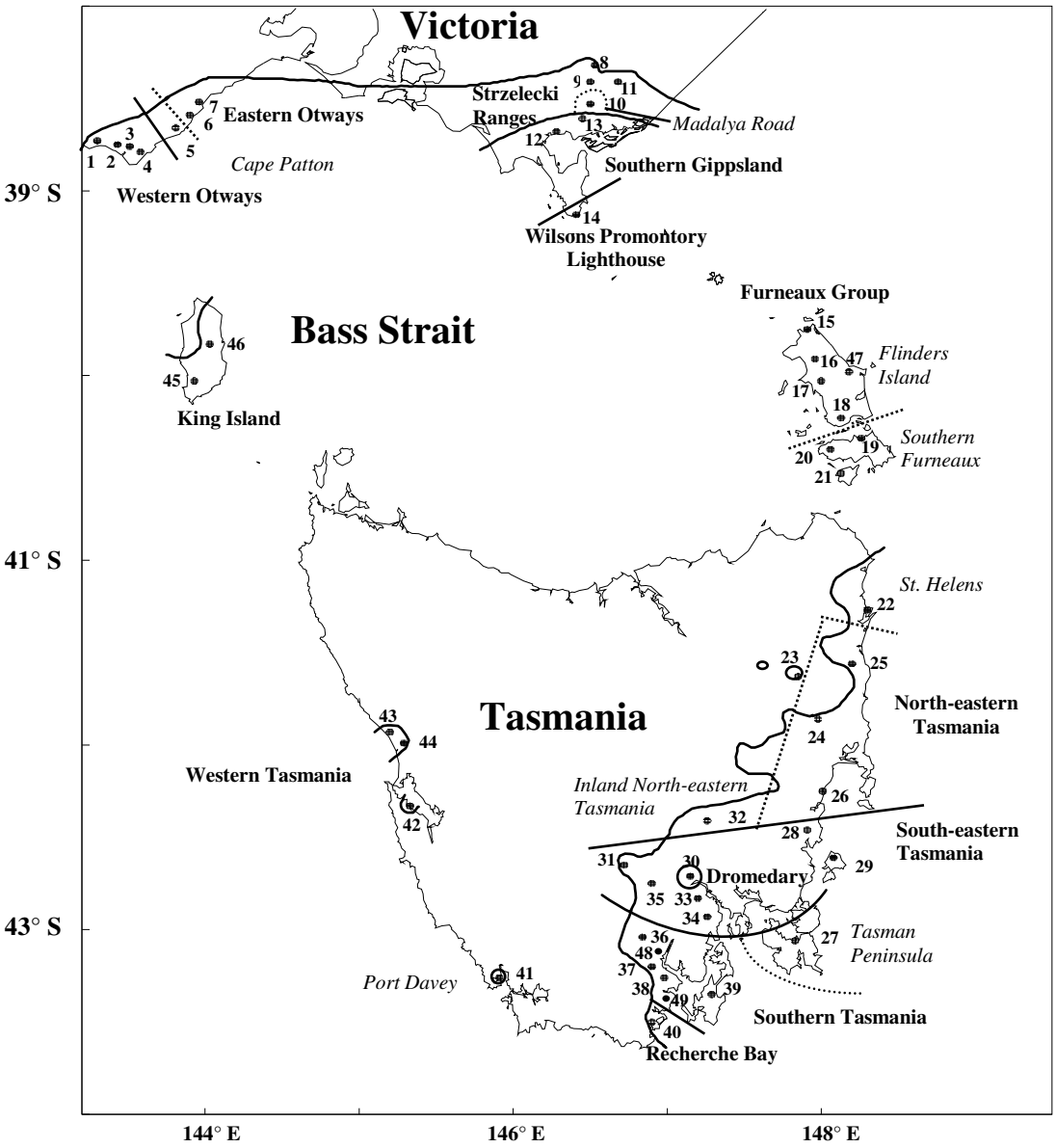


Figure 2-6 The revised racial classification.

The seed collection localities are shown as numbered circles (codes follow Table 2-1). Races are shown in larger type and are separated by solid lines. Sub-races are shown in italics, separated by dotted lines.

2.5.3 Racial Classification

The attributes used in this study and other work on growth and damage by insects and disease of the races for which there was adequate sampling are summarised in Table 2-9. The Western Otways race is characterised by fast growth, high pilodyn penetration, and short distance to the widest point on the leaf, whereas the Eastern Otways have fast growth, low pilodyn penetration and low fungal attack (Carnegie *et al.* 1994), but thick bark and high sawfly damage. The Strzelecki Ranges race is similar to the Eastern Otways race, with fast growth, high leaf width, bark thickness and distance to the widest point on the leaf, but has low pilodyn penetration and high resistance to both drought (Toro *et al.* 1998) and *Phoracantha semipunctata* (Soria and Borralho 1998). The Southern Gippsland race has slow growth, thick bark, low pilodyn penetration, early transition to adult foliage, long distance to the widest point on the leaf and small basal lobes on the leaves. The Wilsons Promontory Lighthouse race is characterised by extremely slow growth, precocious flowering, thick bark, low pilodyn penetration, low drought and *Phoracantha* resistance, early transition to adult foliage and anomalous leaf morphology. All the mainland races are characterised by high defoliation by leaf blister sawfly and autumn gum moth caterpillar (Farrow *et al.* 1994). The Furneaux Group race flowers precociously (especially the southern islands), has high drought and *Phoracantha* resistance and short leaves. The race in north-eastern Tasmania is slow growing with thick bark, long narrow leaves, persistent juvenile foliage and high sawfly damage. The South-eastern Tasmanian race is characterised by delayed flowering and small basal lobes on the leaves, whereas the Southern Tasmania race has thin bark, delayed flowering and long leaves. Western Tasmania has slow to average growth and low *Phoracantha* resistance, but thin bark, early transition to adult foliage and large basal lobes on the leaves. The King Island race is notable for its very thin bark, high pilodyn penetration and high resistance to the three insect species but low resistance to both drought and *Phoracantha*. While King Island showed fast growth in early trials (Volker and Orme 1988), in the present study fast growth was only evident in one of its collection localities (Figure 2-2). The slower growth of the southern locality may, however, be due to selfed seed collected from isolated trees, which has been shown to reduce

growth (Borralho and Potts 1996).

Table 2-8 Allocation of localities to race and sub-race groups.

Race		Sub-Race		Locality	
Code	**Name**	**Code**	**Name**	**Code**	**Name**
1	Western Otways	1	Western Otways	1	South West Lavers Hill
				2	Otway State Forest
				3	Cannan Spur
				4	Parker Spur
2	Eastern Otways	2	Cape Patton	5	Cape Patton
		3	Eastern Otways	6	Jamieson Creek
				7	Lorne
3	Strzelecki Ranges	4	Strzelecki Ranges	8	Jeeralang North
				9	Jeeralang
				11	Bowden Road
		5	Madalya Road	10	Madalya Road
4	Southern Gippsland	6	Southern Gippsland	12	Port Franklin
				13	Hedley
5	Wilsons Promontory Lighthouse	7	Wilsons Promontory Lighthouse	14	Wilsons Promontory Lighthouse
6	Furneaux	8	Flinders Island	15	North Flinders Island
				16	Central North Flinders Island
				47	Central East Flinders Island
				17	Central Flinders Island
				18	South Flinders Island
		9	Southern Furneaux	19	North Cape Barren Island
				20	West Cape Barren Island
				21	Clarke Island
7	North-eastern Tasmania	10	St. Helens	22	St. Helens
		11	North-eastern Tasmania	24	Royal George
				25	German Town
				26	Mayfield
		12	Inland North-eastern Tasmania	23	Pepper Hill
				32	Jericho
8	Dromedary	13	Dromedary	30	Mt. Dromedary
9	South-eastern Tasmania	14	South-eastern Tasmania	28	Triabunna
				29	North Maria Island
				31	Ellendale
				33	Collinsvale
				34	Hobart South
				35	Moogara
10	Southern Tasmania	15	Southern Tasmania	36	Blue Gum Hill
				37	South Geeveston
				38	Dover
				39	South Bruny Island
				48	North Geeveston
				49	Strathblane
		16	Tasman Peninsula	27	Taranna
11	Recherche Bay	17	Recherche Bay	40	Recherche Bay
12	Western Tasmania	18	Port Davey	41	Port Davey
		19	Western Tasmania	42	Macquarie Harbour
				43	Little Henty River
				44	Badgers Creek
13	King Island	20	King Island	45	South King Island
				46	Central King Island

Although the continuous nature of the variation necessitated a hierarchical race and sub-race classification, this analysis has found a far better geographic grouping of localities than analyses based on growth alone (Jordan *et al.* 1994). Although the races identified on growth traits and geographic separation by Jordan *et al.* (1994) are similar to those identified here, this racial classification has relied less on geographic discontinuities in the distribution, and has been able to identify relationships between areas, and subdivisions within areas, that were not apparent in the growth data. For example, we have distinguished between two races in the Otway Ranges, largely on the basis of bark thickness and pilodyn penetration, and between Flinders Island and the southern Furneaux Group on the basis of flowering precocity and juvenile leaf length. Pilodyn penetration, an important trait because of its correlation with wood basic density (Greaves *et al.* 1996), is lower for the Eastern Otway race, which is also more drought tolerant (Dutkowski 1995). Madalya Road (10) has been identified as intermediate between the old (sensu Jordan *et al.* (1994)) Strzelecki Ranges and Southern Gippsland races. The data have also allowed a better discrimination between races in eastern Tasmania. The old North-eastern race has been expanded from just the St. Helens (22) and Pepper Hill (23) localities to cover most of the old Eastern race as well. The old South-eastern race has been divided between the new Southern and South-eastern Tasmanian races. The area north of the Tasman Peninsula has been identified as an area where the three eastern Tasmanian races adjoin. The major trait disjunction identified on growth traits between the Furneaux Group and north-eastern Tasmania, and the affinity between the Furneaux Group and western Tasmania, is supported by this data. Both the present and previous racial classifications indicate a major dichotomy between mainland (including Furneaux Group) and eastern Tasmanian races with the King Island, Western Tasmania and Wilsons Promontory Lighthouse races lying outside, and of undetermined affinity, to these major groups.

Table 2-9 Summary of attributes of major races from our data and other published information.

VL = Very Low, L = Low, A = Average, H = High, VH = Very High.

Race		This Data										Other Data						
Code	Name	Growth	Bark Thickness	Pilodyn Penetration	Flowering Precocity	Leaf length	Leaf Width	Length to Widest Point	Basal Lobing	Height to Phase Change	Sawfly Damage	Growth D	Growth E	Growth F	Insect Defoliation[G]	*Phoracantha*[H] Resistance	Drought[I] Resistance	Fungal[J] Attack
1	Western Otway	H	A	H	A	A	A	L	A	A	A	A			H	A	A	A
2	Eastern Otway	H	H	L	A	A	A	A	A	A	H	H	A			A		L
3	Strzelecki Ra.	H	H	VL	A	A	H	H	A	A	A	H	A		H	H	H	A
4	Southern Gippsland	L	H	L	A	A	A	H	L	L	A					A		
5	Wilsons Promontory Lighthouse	VL[A]	H	L	VH[B]	A[C]	VL[C]	VH[C]	VL[C]	VL[B]		VL			H	L	L	
6	Furneaux	A	A	A	H	L	A	A	A	A	A	A	A	L	A	H	H	
7	NE Tasmania	L	H	A	A	H	L	A	A	H	H	A	A	A	A	A	A	A
9	SE Tasmania	A	A	A	L	A	A	A	L	A	A	A	A	A	A	A		
10	Southern Tasmania	A	L	A	L	H	A	A	A	A	A	A	A	H		A	A	H
12	Western Tasmania	A	L	A	A	A	A	A	H	L	A	A	L	A	A	L	A	
13	King Island	A	VL	VH	A	A	A	A	A	A	L	A	VH	H	L	L	L	A

[A] Height and volume on five sites in Tasmania, Jordan *et al.* (1994)

[B] Hasan (1993)

[C] Potts and Jordan (1994a)

[D] Height growth on five sites in Spain, Vega *et al.* (1994)

[E] Diameter growth on 12 sites in Australia, Kube *et al.* (1995)

[F] Height growth in Portugal, Almeida (1993)

[G] Insect Defoliation by *Mnesampela privata* (Autumn Gum Moth Caterpillar) and *Phylacteophaga froggatti* (Leafblister Sawfly), Farrow *et al.* (1994)

[H] Proportion of trees undamaged by *Phoracantha semipunctata* (Eucalypt long horned stem borer) in Spain, Soria *et al.* (1998)

[I] Survival after drought in Spain, Toro *et al.* (1998)

[J] Severity of fungal attack by *Mycosphaerella* leaf disease - Carnegie *et al.* (1994)

While the localities sampled here fall only into the core *globulus* and *globulus* intergrade populations identified by Jordan *et al.* (1993) on the basis of capsule morphology, the race boundaries are broadly in line with the population boundaries they identified. While capsule morphology is a key taxonomic trait in this complex, much more differentiation has been identified from this analysis. For example, there is a large genetic difference between the Furneaux Group and north-eastern Tasmania which was not reflected in capsule morphology; both areas were mainly of the core *globulus* type. Greater differentiation on our quantitative traits is not always the case. There is no evidence from our data of any difference between the northern Flinders Island population and the rest of the Furneaux Group, despite their more *bicostata*-like capsules (Jordan *et al.* 1993).

The results also show a better grouping of localities than that shown by Nesbitt *et al.* (1995) on the basis of 162 RAPD markers, although both approaches did detect the major difference between the mainland and Tasmanian localities. RAPD markers are a random sample of the genome and most are likely to be adaptively neutral. The data that we have used are likely to provide a more useful classification as the wide variety of traits we used are expressed in the field and, probably being under multi-gene control, they are likely to sample a larger part of the genome.

While this classification is likely to be robust, there are areas of uncertainty because of the variable sampling intensity and poor sampling of some parts of the distribution. This is particularly a problem where there is a potential for steep clines or intergradation with other subspecies. For example, the inland part of the Tasmanian central east coast distribution was only sampled at widely separated locations (Figure 2-1). While these areas have been allocated to the same sub-race, there is the potential for genetic differences to have arisen between these localities, and between them and the coastal populations. Further, the area north of the Tasman Peninsula, which lies at the conjunction of the three east coast Tasmanian races, was not sampled at the time of the collection because of poor seed crops (Gardiner and Crawford 1987, 1988). The poor sampling of these areas makes the delineation of race boundaries difficult. Southern Gippsland, Recherche Bay, Mt. Dromedary and Port Davey are allocated to separate races or sub-races as they are geographic and genetic outliers. However, these areas were not well sampled and so these allocations may change

with better sampling. The clines in variation identified in the Otway Ranges, eastern Tasmania and Gippsland will always cause problems in delineating race boundaries. This problem is exacerbated in the eastern Otways and Gippsland where the populations exist in close proximity with other subspecies, with which there may be a possibility of hybridisation. The Wilsons Promontory Lighthouse (14) population sampled is a cliff-top mallee form, phenotypically similar to populations that occur on exposed sites (not sampled in the collection) in Tasmania at Cape Tourville (Chalmers 1992) and on Maria Island (Brown and Bayly-Stark 1979). Steep clines in a number of traits have been found at Cape Tourville over as little as 2 km from the coast (Chalmers 1992). As this is also likely to be the case at Wilsons Promontory Lighthouse, this locality is unlikely to be representative of the rest of Wilsons Promontory, but where the boundary occurs is unclear from our data.

Despite these limitations, the classification has been shown to be useful in the improved prediction of breeding values (Dutkowski *et al.* 1997), and it also provides a framework for the development of gene conservation strategies for *globulus* and its intergrades. Such strategies are necessary for both *in-situ* and *ex-situ* conservation because, even though the trials of these seedlots used in this study form a world-wide *ex-situ* conservation resource, the collection was not exhaustive and the current management of native forest stands may not recognise the patterns of geographic variation. Priority needs to be given to populations of scattered farmland individuals, as these are most vulnerable because of dieback, lack of regeneration, and clearing.

This work is likely to form the basis of any further classification. The classification will only change in the light of new traits being incorporated, or new localities being sampled. New traits may lead to the further subdivision of current races; however, the range of traits used in this analysis makes this unlikely. Further sampling may be undertaken for a number of reasons, but is unlikely to be of the same size and scope as the one on which this classification is based. Further sampling within the areas of uncertainty identified may well lead to some changes. However, sampling within the range of this classification in areas close to those already sampled is unlikely to lead to changes because of the strong geographic patterns of variation. Such further sampling is only likely to take place if the areas are of economic interest. Sampling in areas further from those sampled here would lead to a geographic extension of the

classification—finding new boundaries or consolidating existing ones. Any new samples would needed to be tested for sufficient traits, and with sufficient geographic overlap, to allow comparison with current races.

2.6 Conclusion

Examination of genetic variation in a wide variety of traits for *Eucalyptus globulus* ssp. *globulus* and its intergrades has shown that there are differences between sample localities, and that the localities can be effectively amalgamated into larger, geographically concordant groups. No simple explanations are available to explain the patterns of variation seen. A revised racial classification has led to the identification of new divisions within races and better identification of race boundaries. Some areas of the distribution may need further sampling to more accurately elucidate their racial affinities, especially those with traits of high economic importance.

2.7 Acknowledgments

The authors acknowledge the contribution of North Forest Products (now Gunns Ltd.) (particularly Ian Ravenwood, Mike Powell and Wayne Tibbits) for establishing and maintaining the trials used in this analysis, and for their permission to collect, and help with collecting, the data. We also thank Andrew MacDonald, Paul Tilyard, Tony Clarke and Peter Gore for help in collection of the data, Andrew MacDonald for the maintenance of the database in which this diverse data is stored, and Greg Jordan for helpful discussion and analysis of the precocity data. We are indebted to Nuno Borralho for his encouragement, discussion and support in carrying out this work.

Chapter 3 Breeding value prediction using the *E. globulus* race classification*

3.1 Summary

The race classification developed for *Eucalyptus globulus* ssp *globulus* was used for the prediction of breeding values for growth, pilodyn penetration, bark thickness and drought susceptibility. Inclusion of sub-races improved the model and reduced the estimates of heritability for traits with large sub-race differences. The distribution of breeding values changed across sub-races and the use of the model which included sub-races increased genetic gains by up to 20%.

3.2 Introduction

Spatial genetic variation within a species can arise where chance or adaptive differentiation cannot be overcome by the homogenising effects of dispersal and migration (Roughgarden 1979). Such is likely to be the case for widespread plant species with no mechanism for widespread dispersal, or for which dispersal barriers exist. *Eucalyptus globulus* Labill. ssp. *globulus*, a native of south eastern Australia, is a species which is likely to exhibit spatial genetic variation. It has a limited migration ability, a widespread distribution, occurs in a number of different environments, and is likely to have had a complex history of environmental change, migration, barriers to dispersal and intergradation with related sub species (Jordan *et al.* 1993; Potts and Jordan 1994a).

---

* Originally published as part of Dutkowski, G.W., Potts, B.M., and Borralho, N.M.G. (1997). Revised racial classification of *Eucalyptus globulus* ssp. *globulus* and the importance of including race in analysis of progeny trials. In proceeding of 'IUFRO Conference on Silviculture and Improvement of Eucalypts'. (Eds AR Higa, E Schaitza, and S Gaiad.) 24-29 August, Salvador, Bahia, Brazil. Vol. 1. pp. 322-329. (EMBRAPA: Colombo, Brazil)

Eucalyptus globulus ssp. *globulus* is also one of the most widespread plantation pulpwood species in the world with over 1,700,00 ha planted, of which 70,000 ha is in Australia (Tibbits *et al.* 1997). There are active breeding programs in Australia (Butcher 1990; Jarvis *et al.* 1995), Chile (Prado and Alvear 1993), Portugal (Borralho and Cotterill 1994), Spain (Vega Alonso *et al.* 1994) and China (Zang *et al.* 1995), almost all of which have a substantial component of early generation trials from a variety of native seed sources.

Regardless of the cause of spatial genetic variation, its existence and identification allows the delineation of geographic racial groups, within which individuals are more similar to each other than to individuals in other groups. Appropriately defined races are of interest to breeders because of the potential to improve the prediction of breeding values, enable links between different seed collections, and standardisation of the classification and nomenclature of seed sources.

Jordan *et al.* (1994) derived a racial classification of the *globulus* subspecies and its intergrades based on growth to age four years in five trials in Tasmania. Although there was clearly spatial variation, much of the variation was continuous and the racial groups were not distinct, so discontinuities in the geographic distribution had to be used to define races. Some localities with few families representing them also proved difficult to classify, despite their geographic proximity to other localities. Data from the same trials used by Jordan *et al.* (1994) for a variety of traits, such as growth, survival, flowering precocity, wood basic density, and leaf morphology, some of which are of economic importance, has been used to revise the classification in to races and sub-races (Dutkowski and Potts 1999) (Chapter 1).

While the use of provenances as fixed effects in forest genetic analyses has been commonplace, the use of an individual additive genetic model (Henderson 1984) for prediction of breeding values in trees is fairly recent. Recent application of such a model in this species has not included any population effect in the estimation of variance components, or the subsequent prediction of breeding values (Jarvis *et al.* 1995). For first generation trials of open-pollinated progeny, the fixed population effect can simply be added to the within population breeding value to obtain a net breeding value for parents and offspring in the trial. More sophisticated approaches

are available for data with mixed breeding populations (Westell *et al.* 1988). This paper reports on the importance of including sub-race in the prediction of breeding values.

3.3 Materials And Methods

The effect of using the race classification was looked at for four traits: DBH, pilodyn and bark thickness from two trials in the data used for the classification (Woolnorth and West Ridgley) , and drought susceptibility data from four trials of the same seedlots from Dutkowski (1995). For each data set, three models were fitted:

NORACE: $y = \mu + Rep + Inblk + Tree + \varepsilon$,

$SUBRACE_f$: $y = \mu + Subrace_f + Rep + Inblk + Tree + \varepsilon$,

$SUBRACE_r$: $y = \mu + Subrace_r + Rep + Inblk + Tree + \varepsilon$

where y is the observation of the variable for the tree, μ is the mean, Subrace is the subrace identified from the racial classification as either a fixed effect ($Subrace_f$) for generation of BLUP's, or a random effect ($Subrace_r$) for testing whether the inclusion of race improves the model using a likelihood ratio test, Rep is the replicate as a fixed effect, Inblk is the incomplete block as a random effect, Tree is the individual tree breeding value as a random effect, and ε is the random error. The net individual tree breeding values were calculated for the $SUBRACE_f$ model by adding the fixed race effect to each tree's breeding value.

The models were compared by the likelihood ratio test (NORACE and $SUBRACE_r$), the estimated heritability, the average breeding values according to the $SUBRACE_f$ model for the top 1 in 40 trees identified from each model, the proportion of selections from the top three races, and the proportion of selections in common. Gain was calculated as the difference between the NORACE and $SUBRACE_f$ models in the average of the breeding values from the $SUBRACE_f$ model. Wilsons Promontory Lighthouse and Port Davey races were not included in the analysis. The models were fitted using the ASReml program (Gilmour *et al.* 1997b).

Table 3-1 The effect of including race into the prediction of breeding values.

The traits analysed were drought susceptibility (DRY, 1-9 scale), diameter (DBH, cm), pilodyn penetration (PILO, mm) and relative bark thickness (BARK, %). P no race effect is the probability of no difference between the $SUBRACE_r$ and NORACE models according to a likelihood ratio test (LRT). Pr is the proportion of total genetic variance due to $Subrace_r$ in the $SUBRACE_r$ model. ΔG is the gain difference between the $SUBRACE_f$ and NORACE models, Pc is the proportion of selections in common, and h^2 is the heritability.

Trait	No. Fam	No. Race	No. Tree	p no Race effect	Pr (%)	ΔG (%)	Pc (%)	Selections from top 3 subraces		h^2	
								SUB-$RACE_f$ Model	NO-RACE Model	SUB-$RACE_f$ Model	NO-RACE Model
DRY	111	10	4175	<0.0001	39	11.3	51	95%	70%	0.19	0.54
	179	8	6668	<0.0001	23	3.8	77	100%	97%	0.12	0.21
	101	8	3584	<0.0001	21	0.3	93	66%	68%	0.16	0.24
	114	6	3948	<0.0001	31	2.8	72	100%	100%	0.19	0.37
DBH	474	19	4293	<0.0001	24	2.8	72	78%	62%	0.23	0.37
	427	19	3943	<0,0001	11	19.7	89	61%	51%	0.33	0.41
PILO	474	19	970	<0.0001	21	19.2	72	46%	24%	0.45	0.61
	427	19	880	<0.0001	22	0.9	77	68%	54%	0.50	0.63
BARK	474	19	970	<0.0001	52	21.6	52	92%	44%	0.13	0.48
	427	19	980	<0.0001	33	8.8	59	100%	68%	0.25	0.61

3.4 Results

The results shown in Table 3-1 indicate that there was a significant improvement between models NORACE and $SUBRACE_r$, i.e. by the inclusion of subrace in the model. In no case were the selections from the two models completely the same. The proportion in common ranged from 51% to 93%. The $SUBRACE_f$ model resulted in gains between 0.3% and 21.6% higher than the NORACE model. Usually more selections came from the three best races, although in the case with the lowest gain slightly more came from the better races with the NORACE model. The differences between the models varied in each case. Generally, as the proportion of the total genetic variance due to the $Subrace_r$ effect increased, the proportion of selections in common decreased, the loss of gain increased, and the difference between the proportion selected from the best sub-races also increased. The heritability according to the NORACE model was substantially higher than with the $SUBRACE_f$ model. In all cases the sum of the $SUBRACE_r$ variance and additive variance was less than the additive variance according to the NORACE model (data not shown). This indicates

that the heritability estimate is biased upwards.

3.5 Discussion

The inclusion of subrace in the estimation of breeding values results in a better model and can substantially change the selections that are made. The difference in selections between the two models is not, however, uniform. The extent of the difference between the models will depend on the strength of the sub-race effect, the distribution of families within races, and the distribution of the race means. As the race effect becomes stronger then, under a race model, the proportion of the breeding value due to race increases, and the proportion of the breeding value due to the phenotypic observation decreases as the heritability decreases. If the sub-race effect is ignored however, the heritability is higher than for the race model and the proportion of the breeding value due to the phenotypic observation is also higher. As the strength of the sub-race effect increases therefore the proportion of common selections should decrease as the emphasis on phenotypic observations, and family and race means changes. As the proportion of families in the best races increases then the likelihood of selections from the best race will increase in both models. However, if the best race has few families, then a high proportion of these are likely to be selected under a race model. The precise distribution of the subrace effects will be important in determining whether there is a difference in the selections made. If there is an outstanding race then the inclusion of a race effect is more likely to result in a change in selections. Whether this is likely to be the case cannot be judged *a priori* for any species, but is more likely to occur as the number of races increases.

Although the incorporation of race is likely to result in better ranking of selections, the result may well be to make more selections from fewer families from a few races. This may result in a quicker development of inbreeding in breeding populations unless specific actions are taken to decrease the relatedness in the selected population. Better breeding values will however mean that the appropriate strategy will be taken consciously, rather than by chance by simply ignoring race in the analysis.

3.6 Conclusion

Incorporation of the *Eucalyptus globulus* spp *globulus* race classification into prediction of breeding values may dramatically alter the selections made. Gains will vary but they can be up to 20% if race effects are large. With a race model the increase in selections from better races will more quickly increase the need to manage inbreeding in the selected population.

3.7 Acknowledgements

The authors would like to acknowledge the permission of North Forest Products (now Gunns Ltd) and Bunnings Treefarms (now Western Australian Plantation Resources) to use the data in this study

Chapter 4 Geographic Genetic Variation in, and Race Classification of, Central Victorian *Eucalyptus nitens*[*]

4.1 Summary

Central Victorian *Eucalyptus nitens* is an important source of genetic material for plantations and breeding programs around the world. A comprehensive and well documented collection of around 400 open-pollinated families (excluding suspected *E. denticulata*) from 28 localities provided the opportunity to examine geographic patterns of quantitative genetic variation in this important region. The collection was well tested, being planted in a series of five trials in Tasmania, each measured for up to 13 traits. Traits included survival, growth, form, bark thickness, transition to adult foliage, pilodyn penetration, flowering incidence and possum damage. Clustering of the localities grouped parts of the Pederick (1979) provenances together: northern Rubicon, northern Toorongo, and most of Macalister (with the exception of Connors Plain, a high altitude sub-provenance) were grouped together, with the southern Rubicon, southern Toorongo, and the new Powelltown provenance forming a separate group. The most distinct locality was Connors Plain, which was separated on adult foliage and bark thickness. The work provides the basis for a racial classification, which should clarify the nomenclature and affinities of seed collections from this area. The proposed race boundaries transgress the traditional provenance boundaries in this area. This collection broke the relationship between fast growth and retention of juvenile foliage observed in trials encompassing the whole of the *E. nitens/E. denticulata* complex.

---

[*]Originaly part published as Dutkowski, G.W., Potts, B.M., Williams, D.R., Kube. P.D. and McArthur, C. (2001). Geographic genetic variation in Central Victorian *Eucalyptus nitens*. In proceedings of IUFRO conference 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. p. 39. 6 pp. (INFOR: Santiago, Chile).

4.2 Introduction

Knowledge of the geographic pattern of genetic variation is one of the essential initial steps in the domestication of any eucalypt species (Eldridge *et al.* 1993). *Eucalyptus nitens* (Deane & Maiden) Maiden is an important plantation species in colder temperate regions of the world (Tibbits *et al.* 1997). Its breeding programs are still in the early stages so knowledge of geographic variation is important. Its native range is in mountainous regions of central and eastern Victoria and eastern New South Wales (Figure 4-1). Since 1991 the Erinundra population in eastern Victoria has been recognised as a separate species – *E. denticulata* (Cook and Ladiges 1991). This population is differentiated by adult leaf and seedling morphology and isozyme variation (Cook and Ladiges 1998). It is recognised that in central Victoria these species' distributions overlap, although they may occupy different habitats (Martin Lavery, *pers. comm.*).

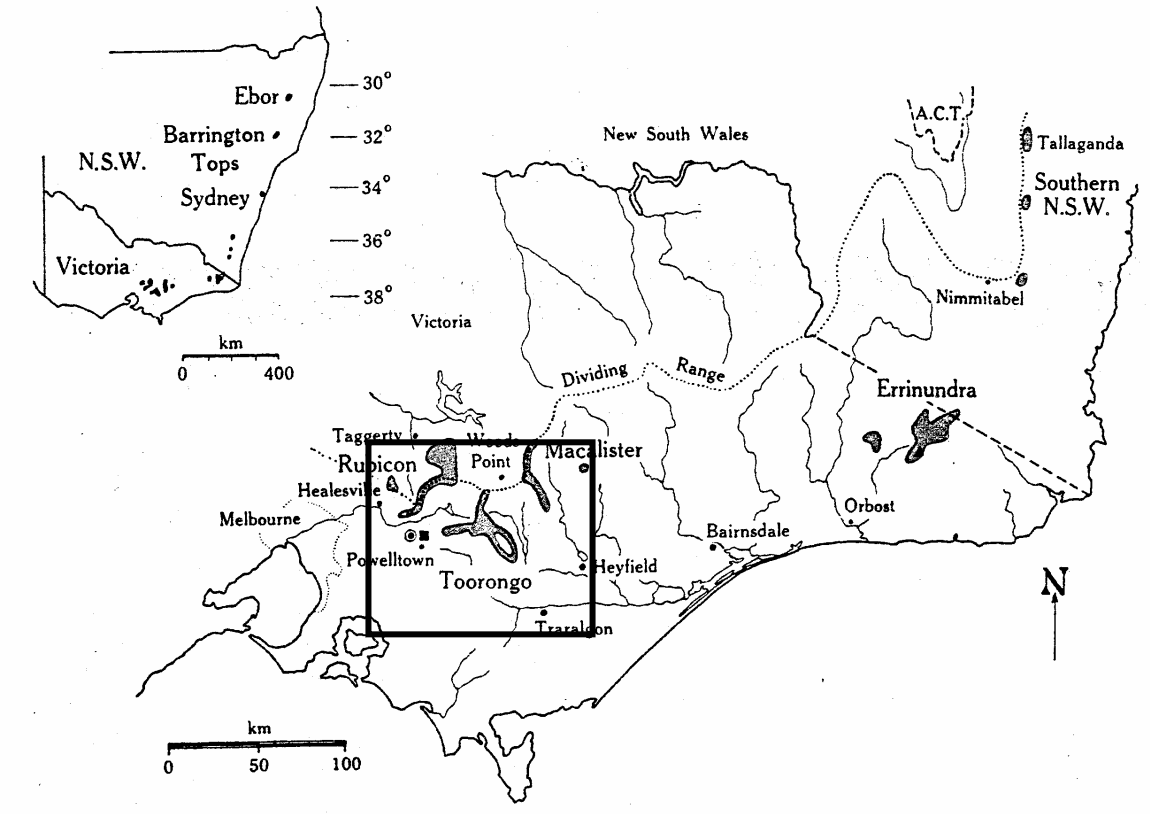


Figure 4-1 Distribution of *Eucalyptus nitens* and *E. denticulata* in Australia.

The inset shows the location of Figure 4-2. From Pederick (1979).

Substantial genetic variation in characters of interest to plantation growers occurs across the species complex, and it has been extensively studied both in Australia and overseas. From the earliest work it has been recognised that *E. denticulata* grew poorly and that the central Victorian provenances were superior in temperate regions (Pederick 1979). Early adult forms from this area, which are presumably *E. denticulata*, also grow poorly. Most temperate breeding programs now concentrate on the central Victorian provenances, although northern NSW provenances have high basic density (Tibbits and Hodge 1998). As most programs are based on early samplings, they contain both species but avoid selecting suspected offspring of *E. denticulata* (Tibbits, *pers. comm.*).

This study is based on a unique and well documented recent collection of the desirable central Victorian provenance. The collection is large and comprehensive in its sampling of the central Victorian distribution, while avoiding any suspected *E. denticulata*. It is well tested and has been measured for a range of traits of interest to tree breeders. The collection enabled us to examine the geographic patterns of genetic variation, and to revise the race classification of this important region.

4.3 Materials and Methods

The collection under study was made in the early 1990's by A. E. O'Connor Pty. Ltd. The collection was well documented and mapped, and areas of suspected *E. denticulata* were avoided. Open pollinated seed was collected from individual parent trees. We grouped the locations of the parent trees into localities of more or less similar geographic size. The localities were generally no more than 5km across and did not cover more than 300m in altitude. Attempts were made to ensure a single physiographic feature was sampled, and that there were sufficient trees sampled in order to obtain a reliable locality mean. All 28 derived localities were geographically separate, except for locality 315 which overlapped with two smaller localities as the precise location of the parent trees was unknown. The localities were grouped into 10 sub-provenances which seem to form more or less continuous areas of distribution. These in turn were grouped into four provenances, three of which followed the names used by Pederick (1979) – Rubicon, Toorongo and Macalister – and a new

provenance from an area to the south unmapped by Pederick that we named Powelltown (Table 4-1).

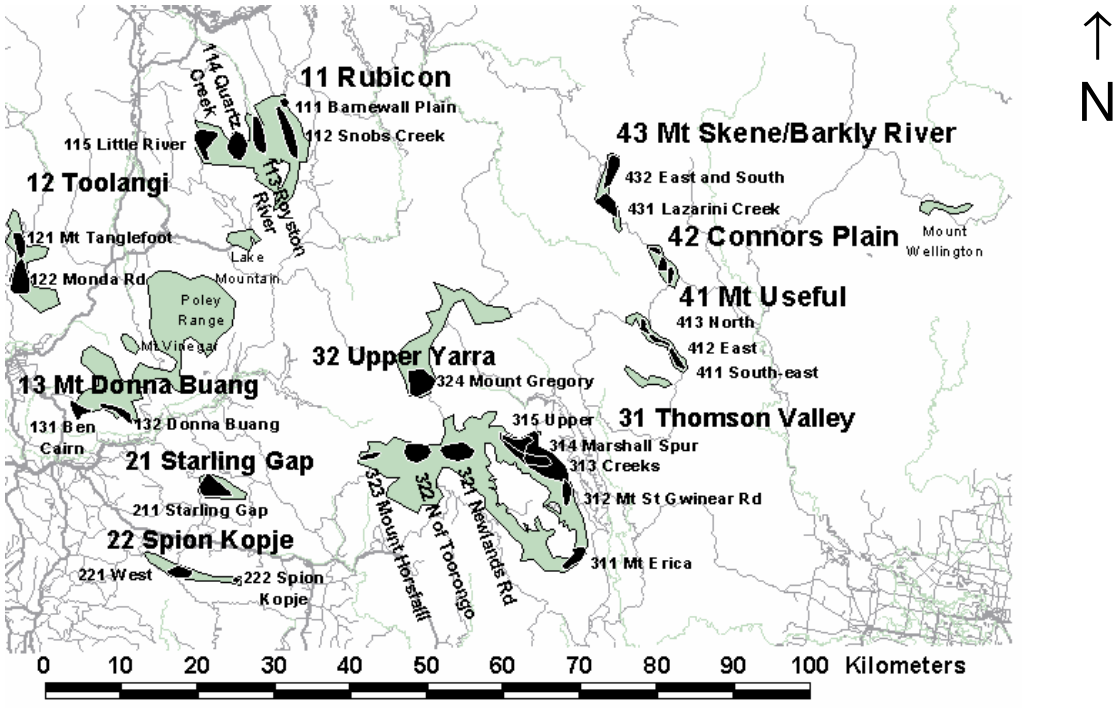


Figure 4-2 Species distribution, sub-provenances, and localities of central Victorian *Eucalyptus nitens*.

Selected other locations named. Background data courtesy of AUSLIG.

The distribution of the species in central Victoria was revised in order to determine the comprehensiveness of the collection that had been undertaken. Distribution records were obtained from four sources: the records of the collection under study, the documented provenance origins from published work on the species, a national herbarium collection database, and the vegetation mapping records of the state forest management authority. These data sets were validated against the each other, the description of the localities, and contour and vegetation maps of the region. Suspect records were excluded. This showed that the collection covered most of the species distribution in the region. Exceptions were an outlying population to the east, at Mt. Wellington, a population to the south of Mt. Useful, and a large area in a restricted access water catchment area centred on the Poley Range (Figure 4-2).

Table 4-1 The geographic origin of localities sampled and families used in the trials.

Provenance	Sub-provenance	Locality	No. Fam's	Latitude (S)	Longitude (E)
1 Rubicon	11 Rubicon	111 Barnewall Plains	15	37° 22´	146° 56´
		113 Royston River	4	37° 24´	146° 53´
		114 Quartz Creek	12	37° 25´	146° 51´
		115 Little River	7	37° 25´	146° 49´
	12 Toolangi	121 Mt. Tanglefoot	27	38° 31´	146° 32´
		122 Monda Road	12	38° 34´	146° 32´
	13 Mt. Donna Buang	131 Ben Cairn	9	38° 43´	146° 37´
		132 Donna Buang	9	38° 43´	146° 41´
2 Powelltown	21 Starling Gap	211 Starling Gap	10	38° 48´	146° 49´
	22 Spion Kopje	221 West	17	38° 54´	146° 46´
		222 Spion Kopje	4	38° 55´	146° 51´
3 Toorongo	31 Thomson Valley	311 Mt. Erica	28	38° 54´	146° 21´
		312 Mt. St. Gwinear Rd	11	38° 49´	146° 20´
		313 Creeks	18	38° 47´	146° 18´
		314 Marshall Spur	9	38° 46´	146° 16´
		315 Upper	32	38° 47´	146° 18´
	32 Upper Yarra	321 Newlands Road	14	38° 46´	146° 11´
		322 North of Toorongo	19	38° 46´	146° 7´
		323 Mount Horsfalll	25	38° 46´	146° 3´
		324 Mount Gregory	37	38° 41´	146° 8´
4 Macalister	41 Mt. Useful	411 South-east	13	38° 40´	147° 30´
		412 East	16	38° 38´	146° 29´
		413 North	11	38° 37´	146° 27´
	42 Connors Plain	421 South-east	12	38° 34´	146° 30´
		422 Plateau	19	38° 33´	146° 29´
		423 North-west	7	38° 32´	146° 28´
	43 Mt. Skene /Barkly River	431 Lazarini Creek	12	37° 29´	146° 24´
		432 East and South	13	37° 27´	146° 25´

Seed from the parent trees were grown in five trials planted in southern Tasmania. Populations north and south of the Poley Range at Lake Mountain and Mt. Vinegar which had been sampled were not tested in the trials. The number of families in each trial varied from 409 to 420, although only 396 were present in all five trials (Table 4-2). The trials were resolvable incomplete block designs, with 5 tree row-plots, 19 or 20 plots per block, and 21 blocks per replicate.

Table 4-2 Distribution of families across the trials and trial design information.

Trial	Code	Families	Replicates	Plots/ Block	Blocks/ Replicate
Florentine	Fl	420	5	20	21
Hollowtree	Ho	407	5	19	21
Taraleah	Ta	417	5	19	21
Meunna	Me	416	5	20	21
Southport	Sp	409	4	20	21
Any Trial		420			
All Trials		396			

The trials were measured up to four times between ages one and six years for a wide variety of traits to yield 45 analysis traits (Table 4-3). Diameter at breast height (DBH, 1.3m) was measured on all trials at age 5 years, from which binary measures of survival and multiple stems were derived. Two of the sites were also assessed for forking, butt sweep and branch size at that age, after having been measured for height at age one. A Forest 6J Pilodyn was used to take two measurements of penetration of a spring-fired pin, as a surrogate for wood density (Greaves *et al.* 1996), in a single window cut into the bark at breast height on a single tree in every plot. It was also used to measure bark thickness at the same place. Precocity was assessed a number of times at three sites as the presence of flowers at an early age. The presence of early adult foliage was assessed as a binary and a proportional trait a number of times at two of the sites, although most often on a sample of trees. Damage to the trees by Brushtail Possums *(Trichosurus vulpecula*) was measured in a variety of ways (including the presence of faeces at the base of the tree!) on half of the Hollowtree

trial.

Table 4-3 Traits measured and their means.

(a) Traits measured on more than one site.

Trait	Age (yrs)	Measure	Site mean				
			Fl	Ho	Ta	Me	Sp
Survival	5	0/1	97%	67%	88%	98%	94%
Height	1	cm				134	122
	3½*	cm	500	452			
DBH	5	cm	10.5	11.0	11.8	11.0	7.3
Multiple Stems	5	0/1	12%	14%	14%	4%	9%
Forks	5	count				0.68	0.49
Butt Sweep	5	0/1				32%	74%
Bark Thickness	5½*	%	9.2%			9.8%	
Internode Length	3½*	cm	22	20			
Branch Size	5	cm class				2.0	1.6
Precocity	3½	0/1	1.5%	1.1%			
	5	0/1	25%	35%			44%
Adult Foliage	3+	0/1		50%			
	3½*	0-1	13%	16%			
	4	0-1	34%				
Pilodyn Penetration	5½*	mm	14.7			15.5	

*=1 tree/plot +=2½ replicates

(b) Possum damage traits measured on 2½ replicates at Hollowtree at age 3 years.

	Damage to Juvenile Foliage	Damage to Adult Foliage	Overall Damage	Broken Branches	Trunk Scratches	Faeces at tree base
Measure	0-4	0-4	0-4	0/1	0/1	0/1
Code	PDJ	PDA	PDO	PBB	PTS	PF
Mean	0.28	0.68	0.49	5.6%	20%	4.6%

The data were extensively validated and cross-validated to ensure consistency between the data sets which had been collected by different groups of researchers. Suspect data points were eliminated, as were missing, dead or runt trees for all traits except survival. Family least squares means were calculated using a linear mixed model including replicate as a fixed effect and block and plot (where appropriate) as random effects. Binomial and proportional data were analysed using a binomial model with logit link function, and count data were analysed using a poisson model with a log link function. All models were fitted using the ASReml software (Gilmour *et al.* 1997b).

Geographic genetic variation was detected by ANOVA of family means within localities using SAS (SAS Institute 1990). Synthetic traits of groups of similar variables were created using principal components analysis with Genstat 5.32 (Payne *et al.* 1988). Locality means of the synthetic traits were used to examine genetic affinities between traits. Correlations were judged to be significantly different from zero on both a single and Bonferroni multiple comparison basis.

Discriminant function analysis (DFA) (Sokal and Rohlf 1981) was used to discriminate between localities. DFA finds linear combinations of family means that maximally differentiate localities in multi-variate space. Locality discriminant scores for all localities were calculated from the discriminant coefficients based on a subset of the data of only those localities with more than three families for all the variables. A hierarchical clustering was undertaken using average linkage clustering of discriminant scores on the significant axes. Race boundaries were proposed from the DFA and hierarchical clustering.

4.4 Results

Localities were significantly different for DBH at all sites, and for height, forks, bark thickness, branch size, adult foliage and pilodyn penetration on the sites at which they were measured (Table 4-4). Survival showed locality differences at Hollowtree only, where it was suspected that frost had caused early mortality. Localities were different in their propensity to have multiple stems at three of the five sites, and for internode length at one of the two sites at which it was measured. Precocity was only

significant at the later age, when the proportion of flowering trees rose above 2%. Of the five possum damage traits, localities were significantly different for three traits.

Table 4-4 Probability that there are no locality mean differences.

(a) Traits measured on more than one site.

Trait	Age	Site				
	(years)	Fl	Ho	Ta	Me	Sp
Survival	5	0.635	0.001	0.210	0.127	0.910
Height	1	<0.001	<0.001		<0.001	<0.001
	3½					
DBH	5	0.001	0.030	<0.001	<0.001	0.001
Multiple Stems	5	0.468	0.661	0.014	0.059	0.006
Forks	5				0.005	0.001
Butt Sweep	5				<0.001	0.072
Bark Thickness	5½	<0.001			0.001	
Internode Length	3½	0.003	0.091			
Branch Size	5				<0.001	0.035
Precocity	3½	0.632	0.026			<0.001
	5	0.795	0.012			
Adult Foliage	3		<0.001			
	3½	<0.001	<0.001			
	4	<0.001				
Pilodyn Penetration	5½	0. 001			<0.001	

(b) Possum damage traits at Hollowtree.

Damage to Juvenile Foliage	Damage to Adult Foliage	Overall Damage	Broken Branches	Trunk Scratches	Faeces at tree base
<0.001	0.027	0.155	0.263	0.021	0.109

The strongest correlation between the locality synthetic traits was between adult foliage and bark thickness (-0.75, Table 4-5). There were a number of other significant correlations as well, mostly involving adult foliage, bark thickness, growth and internode length. There was a strong positive correlation between adult foliage and growth.

Table 4-5 Correlations between locality means of synthetic traits.

Significantly different from 0: single test: $^{+}$ <0.05, $^{++}$ <0.01, Bonferroni test: *<0.05, ** <0.01.

	Survival	Multi-Stems	Butt Sweep	Growth	Precocity	Adult Foliage	Pilodyn Penetration	Bark Thickness	Branch Size
Multi-Stems	0.35								
Butt Sweep	$^{+}$0.45	0.12							
Growth	-0.22	-0.36	-0.23						
Precocity	-0.01	-0.35	-0.11	$^{+}$0.45					
Adult Foliage	-0.35	-0.25	$^{+}$-0.53	*0.63	$^{+}$0.49				
Pilodyn Penetration	0.33	*0.59	0.03	-0.13	-0.36	0.00			
Bark Thickness	0.12	0.06	$^{+}$0.53	$^{++}$-0.57	-0.04	**-0.75	-0.35		
Branch Size	0.01	0.10	0.32	0.12	0.23	0.06	0.05	0.16	
Internode Length	-0.29	0.04	$^{+}$-0.48	**0.65	0.16	*0.62	0.09	*-0.63	0.03

The first six discriminant axes were significant, explaining 70% of the family variation. Plotting the discriminant scores on the first three discriminant axes showed that there was significant geographic structure to the variation (Figure 4-3). The discriminant coefficients and the ANOVA *F*-ratio showed that adult foliage and bark thickness were making a large contribution to that structure. Most localities clustered within their sub-provenance, although Barnewall Plain (111) from Rubicon and Mt. Gregory (324) from Upper Yarra were closer to the Macalister sub-provenances of Mt.Skene/Barkly River and Mt. Useful. There was a cline in variation in the

Thomson Valley and the sub-provenance of Connors Plain was distinctly different from the surrounding Macalister sub-provenances.

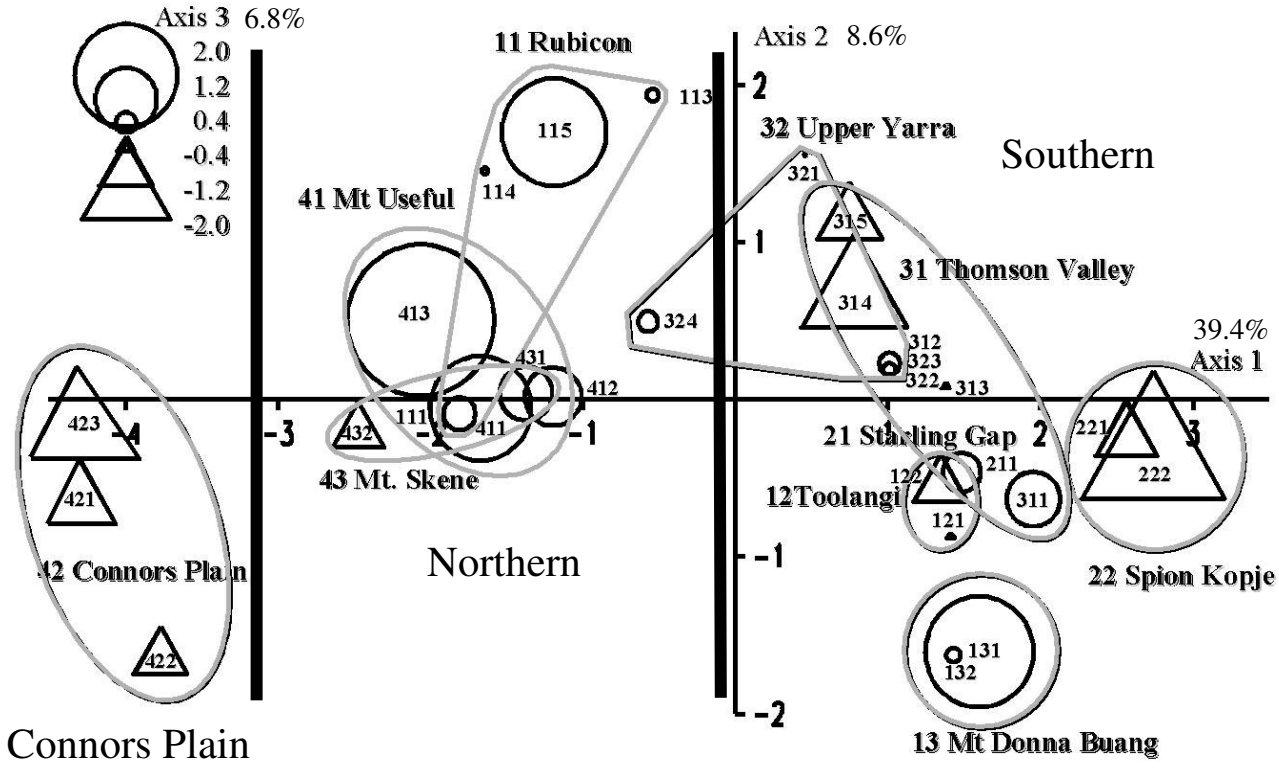


Figure 4-3 Discriminant function scores for localities on the first three axes.

Localities are numbered, sub-provenance groups have grey lines and are numbered, and the proposed races are separated by black lines.

The hierarchical clustering confirmed that there was a major north-south disjunction, and that Connors Plain formed an outlying group (Figure 4-3). The boundary between the north and south put most of Pederick's original Macalister provenance (except Connors Plain) together with the northern part of his Rubicon provenance and the northernmost locality (Mt. Gregory-324) of the Toorongo provenance. The sub-provenances of Toolangi and Mt. Donna Buang clustered with the southern group, not with their Rubicon provenance. Connors Plain, although geographically located between the other Macalister sub-provenances, was quite different.

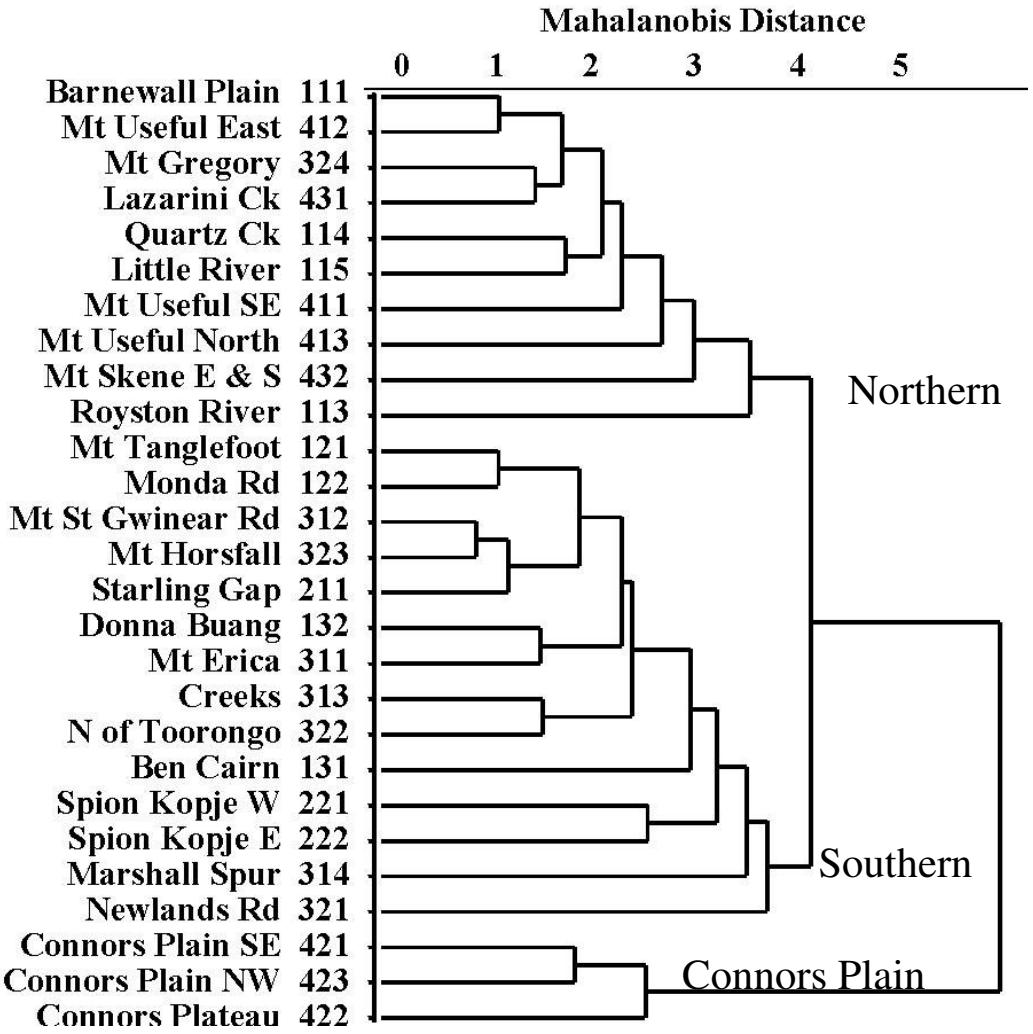


Figure 4-4 Average linkage clustering of the significant discriminant axis scores and the proposed races.

On the basis of these results we propose dividing central Victorian *E. nitens* into three races, differentiated at a Mahalanobis distance of 3.8. These three races we have called Northern, Southern, and Connors Plain.

4.5 Discussion

The races proposed here should be robust as they are based on a very thorough sampling of the central Victorian *E. nitens*, and a comprehensive series of measurements of a wide variety of traits.

While the races are geographically distinct, there is no obvious physiographic reason for the differences, other than the high altitude of Connors Plain. While the Southern

race tends to fall on the southern slopes of the Great Dividing Range, in a number of places the race boundary transgresses the ridge. It brings together some sub-provenances which are separated by large distances, while separating some sub-provenances that are quite close. The aberrant nature of the Connors Plain race has been noted by others (Pederick 1979), but this study indicates that it is different enough to require a separate race.

The races that we have identified cut across the previous provenance divisions for the region and these divisions should be abandoned. Means based on the old provenance will be less likely to show significant differences as they compound genetically distinct races. Trial data should be reanalysed using these boundaries to better understand the patterns of geographic variation.

While the collection used was the most comprehensive one of this region, there are still areas that were not collected and, because they lie on race boundaries, their affinities are still unknown. This includes the large unsampled area between the Mt. Donna Buang and Rubicon sub-provenances. Although the southern part of the distribution in this area seems to be contiguous with Mt. Donna Buang, it also contains the disjunct Lake Mountain and Mt. Vinegar populations, and the location of the boundary remains unknown.

Interestingly, our study showed a reversal of the correlation found by Pederick (1979) as we found adult foliage was correlated with faster growth. However his trials included *E. denticulata*, which has both slow growth and early adult foliage.

4.6 Conclusion

A new race classification of Central Victorian *Eucalyptus nitens* has been proposed – we commend its use to those interested in geographic genetic variation of this region.

4.7 Acknowledgements

The authors would like to thank Norske Skog, Forestry Tasmania, and the Southern Tree Breeding Association for the permission to use the tree performance data used in this study, Environment Australia for the national herbarium record database, and the

Victorian Department of Natural Resources and Environment for their vegetation records. Most significantly we wish to thank Chris O'Connor and Martin Lavery for their excellent seed collection and record keeping, without which this study would have been impossible.

Chapter 5 Inversion of the A matrix under parental inbreeding and partial selfing*

5.1 Summary

Partial selfing and parental inbreeding is a common feature of a number of tree genera, including eucalypts. Analysis of forest genetics trials using an individual tree model requires the correct specification of the additive relationship between trees, and with their parents, to give unbiased estimates of variance components and breeding values. Partial selfing in open pollinated families increases the parent-offspring and offspring-offspring relationships and leads to increased inbreeding. Failure to account for this leads to upwardly biased estimates of additive variance. A modified among offspring relationship is commonly used with a family model to estimate variance components. An algorithm was developed to modify the additive relationship matrix, and generating its inverse using simple rules, where parental inbreeding and partial selfing occurs. The algorithm has been implemented in a FORTRAN program to generate files to substitute for the inverse of the additive relationship matrix normally used in the ASReml program. Use of the modified relationship matrix corrected the inflation in heritability estimates in simulated data sets. Using the inflated heritability and incorrect relationship matrix leads to inflated breeding value estimates. If the correct variance components were used with an incorrect relationship matrix, then the correlation between breeding values was high, but the offspring breeding values were deflated and parental breeding values were inflated.

5.2 Introduction

The use of an individual model for the prediction of breeding values using Best Linear Unbiased Prediction (BLUP) is becoming increasingly popular in tree breeding (Jarvis

---

* Part published as Dutkowski, G.W. and Gilmour, A.R. (2001). Modification of the additive relationship matrix for open pollinated trials. In proceeding of IUFRO conference 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. p. 71. 6 pp. (INFOR: Santiago, Chile).

et al. 1995; Araújo *et al.* 1997; Fernandez *et al.* 1998; Soria *et al.* 1998; Wei and Borralho 2000; Apiolaza and Garrick 2001).

The standard mixed model equation used in such analyses is

[5-1] $$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where **y** is the vector of data, **b** is a vector of fixed effects with its design matrix **X**, **u** is a vector of random effects with its design matrix **Z**, and **e** is a vector of residuals. Fixed effects factors may include the mean, site, replicate and genetic group effects, and the random effects factors may include incomplete block and plot effects as well as additive genetic effects (breeding values). Solutions are obtained by solving the mixed model equations (MME) (Henderson 1984).

[5-2] $$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

where **R** is the variance/covariance matrix of the residuals and where the random effect terms are assumed to be independent, **G** is the direct sum of the variance/covariance matrices of each of the random effects.

Usually the levels within each random effects factor are assumed to be independent, so that the variance/covariance matrices in **G** are identity matrices scaled by the appropriate variance. When the levels are not independent a relationship matrix between the effects replaces the identity matrix. For breeding values the additive (or numerator) relationship matrix (**A**, or NRM) is used. It can accommodate all the trees measured across generations, as well as their unmeasured parents and relatives (Henderson 1976), subject to the assumption that the additive variance is the same for all of the populations from which the founder parents came. For multivariate analysis, the **G** sub-matrices are direct products of the inter-trait variance/covariance matrices and the appropriate relationship matrices.

Clearly the MME's are very flexible as they can be extended to a large number of traits and observations, and fixed and random effects, while allowing for relationships between levels and using the relationship between traits. For tree breeders this means

that they can potentially include all the information in their programs in a single analysis, as long as they know the relationship between the trees, and the variances and covariances of the random effects. There are a number of impediments to the implementation of the MME's in tree breeding (White and Hodge 1989), but analytical approaches and matching software are being developed to overcome these problems (Kerr *et al.* 2001).

Critical to the application of BLUP is the use of the numerator relationship matrix (NRM, usually denoted as **A**). It contains the additive relationship between all trees (a_{ij}) , measured or unmeasured, and can include all parents and their offspring to allow simultaneous prediction of breeding values for both. The NRM has the following properties (Henderson 1976):

1. Symmetry

2. $a_{ii}=1+f$, where f is Wright's inbreeding coefficient.

3. $a_{ij}=r_{ij}\sqrt{(a_{ii}a_{jj})}$, where r is Wright's coefficient of relationship

To construct **A** using Wright's path coefficient method (Wright 1934) is cumbersome, but a recursive method has been developed to construct it easily (Henderson 1976):

1. Order progeny below parents

2. Proceed along each row of the matrix

3. For base parents, which are assumed to be non-inbred and unrelated, A is an identity matrix

4. Between parents and offspring, and between offspring the relationship is $a_{ij} = a_{ji} = 0.5(a_{i\text{-dam}} + a_{i\text{-sire}})$

5. The inbreeding of offspring is $a_{ii} = 1 + 0.5a_{\text{dam-sire}}$

The inverse of the relationship matrix is required in the MMEs and simple rules have been developed to allow its rapid formation without actually forming the relationship matrix itself (Henderson 1976; Quaas 1978). These rules have been extended to allow

for uncertainty in the identity of the parents (Henderson 1987; Famula 1992) and to include fixed group effects in the breeding values (Westell *et al.* 1988).

Under the assumption that the base parents are non-inbred and unrelated (Henderson 1976) derived rules for the inversion of **A** using a cholesky decomposition such that **A** = **LL**', where **L** = **TD** where **T** is a lower triangular matrix that describes the transfer of genes from one generation to the next and **D** is the diagonal elements of **L**. The inverse of **A** required for the solution to the MME can be calculated as

[5-3] $\mathbf{A}^{-1} = (\mathbf{T}^{-1})'(\mathbf{D}^{-1})^2\mathbf{T}^{-1}$

for which only the inverses of **T** and **D** are required. The inverse of **T** can be derived as (**I**-**P**) where **I** is an identity matrix and **P** describes the gametic contribution of parents (1 on diagonals, 0.5 for parent-offspring relations [-0.5 in inverse], 0 otherwise) (Kennedy 1989). From the equality a=0.5**P**a + h which relates the breeding value of offspring to parents, where in this case **P** is the design matrix of parent offspring relationships, Famula (1992) derived **T** as **I**-0.5**P**,which is the same result but with a slight different definition of **P**.

Among animals there are often families with a known father but unknown mother. This arises when a sire is introduced into a field of dams and offspring performance is recorded but maternity is not. This is analogous to open pollinated families in tree breeding trials, but the identity of the mother from which the seed was collected is known, and the father is not. This is simply handled in construction of a NRM and the matrix in both situations is the same: the parent-offspring additive relationship is 0.5, and that between offspring is 0.25. However, base generation open pollinated tree breeding progeny trials often fail to meet the assumptions made in the construction of the usual **A** matrix. The base parents trees may be inbred and related to each other, and the open pollinated offspring will include matings with related trees, including selfs in species with mixed mating systems. Many of these features of open pollinated families will increase the among offspring relationship. This will bias estimates of additive genetic variance if a coefficient of relationship of 0.25 is applied to the family variance (Squillace 1974; Askew and El-Kassaby 1994). Estimates of additive variance often therefore use some other coefficient of relationship to estimate the additive variance, and Hodge *et al.* (1996) report that values between 0.25 and

0.54 have been used. This variation creates problems in the comparison of heritability estimates across different studies (Lopez *et al.* 2002).

While the effect of open pollination on variance component estimation has been studied, there has been little work on its effect on individual breeding value estimates using the NRM. In applying an individual model in tree breeding, Jarvis *et al.* (1995) simply used a corrected estimate of additive variance from a family model, but did not attempt to modify the NRM to account for the increased level of relatedness. Soria *et al.* (1998) modified the NRM using the rules for its creation and inversion under parental uncertainty (Perez-Enciso and Fernando 1992), where the female parent was assigned a probability of being the male parent as well, to both estimate variance components and predict breeding values. However they did not compare this with any other approach.

In this study we have used parental inbreeding and partial selfing as surrogates for all the effects which increase the relatedness in open pollinated seed. We show how this increases both the parent-offspring and among-offspring relatedness, and derive simple rules for inversion of the NRM under different levels of selfing and parental inbreeding. Using these rules we have examined the effect of incorrect assumptions about relatedness on the estimation of variance components and breeding value predictions for both parents and offspring in a simulated open pollinated progeny trial.

5.3 Materials and Methods

5.3.1 Derivation of A^{-1} under partial selfing and parental inbreeding

Consider the situation of an open pollinated family with a partially inbred parent and some proportion of offspring in the trials planted from that family being the result of selfing. By using the tabular method (Henderson 1976) we can compute the average NRM for a parent and its offspring.

Initially, consider a parent (P) with a degree of inbreeding ($a_{pp} = 1+F$, where F is Wright's inbreeding coefficient) which has four offspring; two selfed (S_1 & S_2) with parents P & P, and two outcrossed (O_1 & O_2), with parents P & U (Unknown and unrelated). Applying the tabular method gives the following NRM, **A**:

[5-4]

$$\mathbf{A} = \begin{array}{cc} & \begin{array}{ccccc} \text{UU} & \text{PP} & \text{PP} & \text{PU} & \text{PU} \\ \text{P} & \text{S}_1 & \text{S}_2 & \text{O}_1 & \text{O}_2 \end{array} \\ \begin{array}{cc} \text{UU} & \text{P} \\ \text{PP} & \text{S}_1 \\ \text{PP} & \text{S}_2 \\ \text{PU} & \text{O}_1 \\ \text{PU} & \text{O}_2 \end{array} & \left[\begin{array}{c|cc|cc} a_{PP} & \frac{a_{PP}+a_{PP}}{2}=a_{PP} & a_{PP} & \frac{a_{PP}+a_{PU}}{2}=\frac{a_{PP}}{2} & \frac{a_{PP}}{2} \\ \hline a_{PP} & 1+\frac{a_{PP}}{2} & \frac{a_{S_1P}+a_{S_1P}}{2}=a_{PP} & \frac{a_{S_1P}+a_{S_1U}}{2}=\frac{a_{PP}}{2} & \frac{a_{PP}}{2} \\ a_{PP} & a_{PP} & 1+\frac{a_{PP}}{2} & \frac{a_{S_2P}+a_{S_2U}}{2}=\frac{a_{PP}}{2} & \frac{a_{PP}}{2} \\ \hdashline \frac{a_{PP}}{2} & \frac{a_{PP}}{2} & \frac{a_{PP}}{2} & 1+\frac{a_{PU}}{2}=1 & \frac{a_{O_1P}+a_{O_1U}}{2}=\frac{a_{PP}}{4} \\ \frac{a_{PP}}{2} & \frac{a_{PP}}{2} & \frac{a_{PP}}{2} & \frac{a_{PP}}{4} & 1 \end{array}\right] \end{array}$$

Now consider the average NRM ($\mathbf{A}^s$), where the offspring (O^s_1 & O^s_2) of the parent (P) have a probability (*s*) of being selfed and (*1-s*) of being outcrossed:

[5-5]

$$\mathbf{A}^s = \begin{array}{cc} & \begin{array}{ccc} & \text{s}(\text{S}_1) : \text{PP} & \text{s}(\text{S}_2) : \text{PP} \\ \text{UU} & (1-\text{s})(\text{O}_1) : \text{PU} & (1-\text{s})(\text{O}_2) : \text{PU} \end{array} \\ \begin{array}{cc} \text{UU} & \text{P} \\ \text{s}(\text{S}_1) : \text{PP} & \\ (1\text{-s})(\text{O}_1):\text{PU} & \text{O}^s_1 \\ \text{s}(\text{S}_2):\text{PP} & \\ (1-\text{s})(\text{O}_2) : \text{PU} & \text{O}^s_2 \end{array} & \left[\begin{array}{c|cc} a^s_{PP} & a^s_{PO_1} & a^s_{PO_2} \\ \hline a^s_{O_1P} & a^s_{O_1O_1} & a^s_{O_1O_2} \\ a^s_{O_2P} & a^s_{O_2O_1} & a^s_{O_2O_2} \end{array}\right] \end{array}$$

Averaging the relationships in [5.4] according to their probabilities gives:

[5-6] $a^s_{PP} = a_{PP}$

[5-7]

$$\begin{aligned} a^s_{PO_1} &= sa_{PS_1} + (1-s)a_{PO_1} = sa_{pp} + (1-s)a_{pp}/2 = a_{pp}\frac{(s+1)}{2} \\ &= a^s_{PO_2} = a^s_{O_1P} = a^s_{O_2P} \end{aligned}$$

[5-8]

$$\begin{aligned} a^s_{O_1O_1} &= sa_{S_1S_1} + (1-s)a_{O_1O_1} = s(1+a_{PP}) + (1-s)1 \\ &= s + \frac{sa_{PP}}{2} + 1 = \frac{sa_{PP}+2}{2} = a^s_{O_2O_2} \end{aligned}$$

[5-9]

$$\begin{aligned} a^s_{O_1O_2} &= s^2 a_{S_1S_2} + s(1-s)a_{S_1O_2} + (1-s)sa_{O_1S_2} + (1-s)^2 a_{O_1O_2} \\ &= s^2 a_{PP} + s(1-s)a_{PP} + (1-s)^2 a_{PP}/4 \\ &= s^2 a_{PP} + sa_{PP} - s^2 a_{PP} + a_{PP}/4 - 2sa_{PP}/4 + s^2 a_{PP}/4 \\ &= \tfrac{a_{PP}}{4}(4s + 1 - 2s + s^2) = \tfrac{a_{PP}}{4}(1+s)^2 = a_{PP}\left(\tfrac{1+s}{2}\right)^2 \\ &= a^s_{O_2O_1} \end{aligned}$$

Then substituting [5-6] to [5-9] into [5-5] gives:

$$\textbf{[5-10]}\quad \mathbf{A}^s = \begin{array}{c} \\ \mathrm{P} \\ \mathrm{O}_1^s \\ \mathrm{O}_2^s \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{P} & \mathrm{O}_1^s & \mathrm{O}_2^s \end{array} \\ \begin{bmatrix} a_{PP} & a_{PP}\frac{(s+1)}{2} & a_{PP}\frac{(s+1)}{2} \\ a_{PP}\frac{(s+1)}{2} & \frac{sa_{PP}+2}{2} & a_{PP}\left(\frac{1+s}{2}\right)^2 \\ a_{PP}\frac{(s+1)}{2} & a_{PP}\left(\frac{1+s}{2}\right)^2 & \frac{sa_{PP}+2}{2} \end{bmatrix} \end{array}$$

From [5-10] we can see that the parental inbreeding rate and the selfing rate increase the parent-offspring relationship (Figure 5-1), the offspring-offspring relationship (Figure 5-2), and the offspring inbreeding (Figure 5-3).

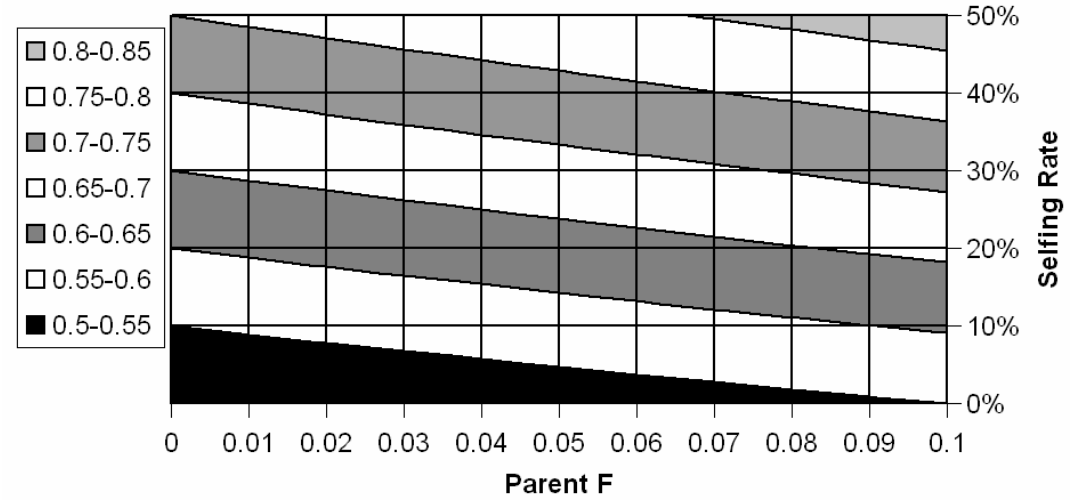


Figure 5-1 The effect of parental inbreeding (F) and selfing rate on the parent-offspring relationship.

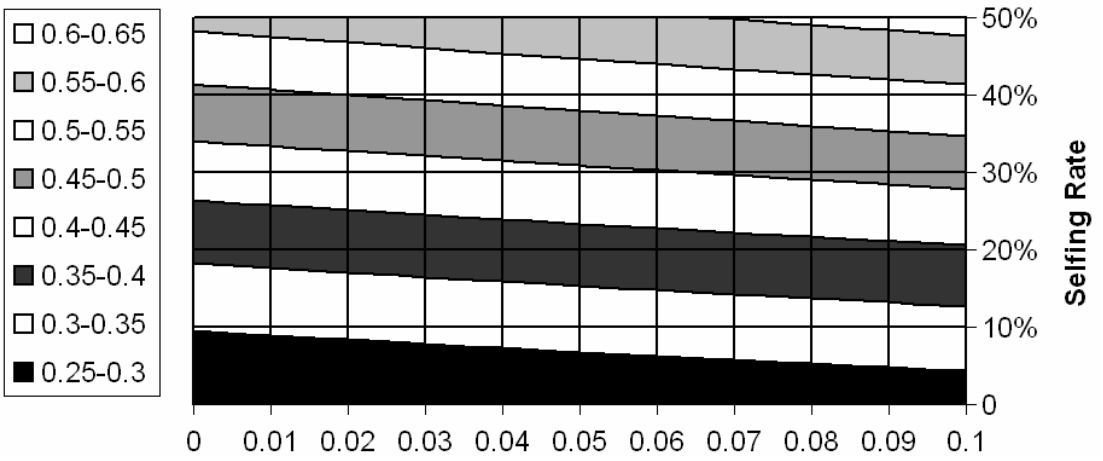


Figure 5-2 The effect of parental inbreeding (F) and selfing rate on the offspring-offspring relationship.

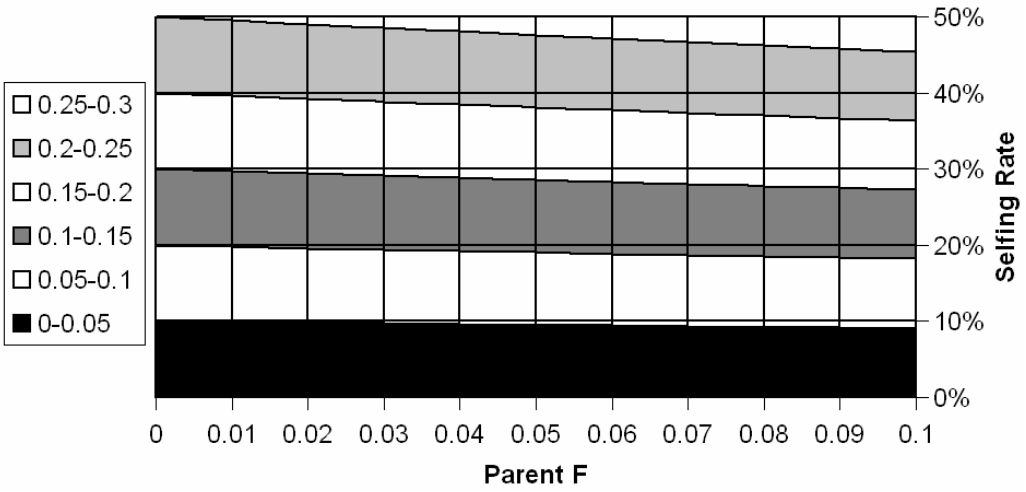


Figure 5-3 The effect of parental inbreeding (F) and selfing rate on the offspring inbreeding (F).

This $\mathbf{A}^s$ NRM for 1 parent and n offspring has the following structure:

$$\mathbf{A}^s = \begin{array}{c} \\ \text{P} \\ \text{O} \end{array} \begin{array}{c} \begin{array}{cc} \text{P} & \text{O} \end{array} \\ \begin{bmatrix} a & ab\mathbf{1}_n \\ ab\mathbf{1}_n' & ab^2\mathbf{J}_n + (c - ab^2)\mathbf{I}_n \end{bmatrix} \end{array} \quad \textbf{[5-11]}$$

where

$$a = a_{PP} \quad \textbf{[5-12]}$$

$$b = \frac{s+1}{2} \quad \textbf{[5-13]}$$

$$c = \frac{sa_{PP} + 2}{2} \quad \textbf{[5-14]}$$

and $\mathbf{I}_n$ is an identity matrix of size n, $\mathbf{J}_n$ is a square matrix of 1's of size n, and $\mathbf{1}_n$ is a vector of 1's of size n.

Consider that now $\mathbf{A}=\mathbf{A}^s$ is the matrix of interest. Inversion of this NRM proceeds by decomposition into a lower triangular matrix **L**, and a diagonal matrix **D**, such that

[5-15] A = LDL'

In order to determine the structure of these matrices we consider a 4x4 matrix:

$$\mathbf{A} = \begin{bmatrix} a & ab & ab & ab \\ ab & c & abb & abb \\ ab & abb & c & abb \\ ab & abb & abb & c \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ d & 1 & 0 & 0 \\ e & f & 1 & 0 \\ g & h & i & 1 \end{bmatrix} \begin{bmatrix} j & 0 & 0 & 0 \\ 0 & k & 0 & 0 \\ 0 & 0 & l & 0 \\ 0 & 0 & 0 & m \end{bmatrix} \begin{bmatrix} 1 & d & e & g \\ 0 & 1 & f & h \\ 0 & 0 & 1 & i \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

[5-16]

$$= \begin{bmatrix} j & 0 & 0 & 0 \\ dj & k & 0 & 0 \\ ej & fk & l & 0 \\ gj & hk & il & m \end{bmatrix} \begin{bmatrix} 1 & d & e & g \\ 0 & 1 & f & h \\ 0 & 0 & 1 & i \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} j & jd & je & jg \\ dj & ddj + k & dej + fk & dgj + kh \\ ej & dej + fk & eej + ffk + l & ejg + fkh + li \\ gj & djg + hk & gej + hkf + il & ggj + hhk + iil + m \end{bmatrix}$$

So that the following equalities hold:

[5-17]

$$\begin{aligned} & j = a \\ & ab = jd = je = jg = ad = ae = ag \Rightarrow d = e = g = b \\ & c = ddj + k && \Rightarrow k = c - ddj = c - abb \\ & abb = dej + fk && \Rightarrow f = (abb - dej)/k = 0 \\ & abb = djg + hk && \Rightarrow h = (abb - djg)/k = 0 \\ & c = eej + ffk + l && \Rightarrow l = c - eej - ffk = c - abb \\ & abb = gej + hkf + il && \Rightarrow i = (abb - gej - hkf)/(c - abb) = 0 \\ & c = ggj + hhk + iil + m && \Rightarrow m = c - ggj - hhk - iil = c - abb \end{aligned}$$

Thus, the structures of **L** and **D** in this example, and more generally, are:

[5-18]

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ b & 1 & 0 & 0 \\ b & 0 & 1 & 0 \\ b & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}_n \\ b\mathbf{1}_n' & \mathbf{I}_n \end{bmatrix}$$

where $\mathbf{0}_n$ is a vector of 0's of size n, and

[5-19]

$$\mathbf{D} = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & c - abb & 0 & 0 \\ 0 & 0 & c - abb & 0 \\ 0 & 0 & 0 & c - abb \end{bmatrix} = \begin{bmatrix} a & \mathbf{0}_n \\ \mathbf{0}_n' & (c - abb)\mathbf{I}_n \end{bmatrix}$$

The inverse of **A** is

[5-20] $\mathbf{A}^{-1} = \left(\mathbf{L}^{-1}\right)' \mathbf{D}^{-1}\mathbf{L}^{-1}$

for which the inverses of **L** and **D** are required:

[5-21] $\mathbf{L}^{-1} = \begin{bmatrix} 1 & 0_{\mathbf{n}} \\ -b\mathbf{1}_n' & \mathbf{I}_n \end{bmatrix}$

which can be shown from

$$\mathbf{I} = \mathbf{L}^{-1}\mathbf{L} = \begin{bmatrix} 1 & 0_{\mathbf{n}} \\ -b\mathbf{1}_n' & \mathbf{I}_n \end{bmatrix}\begin{bmatrix} 1 & 0_{\mathbf{n}} \\ b\mathbf{1}_n' & \mathbf{I}_n \end{bmatrix}$$

[5-22]
$$= \begin{bmatrix} 1+0_{\mathbf{n}}b\mathbf{1}_n' & 10_{\mathbf{n}}+0_{\mathbf{n}}\mathbf{I}_n \\ -b\mathbf{1}_n'1+\mathbf{I}_n b\mathbf{1}_n' & -b\mathbf{1}_n'0_{\mathbf{n}}+\mathbf{I}_n\mathbf{I}_n \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0_{\mathbf{n}} \\ 0_{\mathbf{n}} & \mathbf{I}_n \end{bmatrix} = \mathbf{I}$$

and the inverse of a diagonal matrix is the inverse of each of the diagonal elements:

[5-23] $\mathbf{D}^{-1} = \begin{bmatrix} \frac{1}{a} & \mathbf{0}_n \\ \mathbf{0}_n' & \frac{1}{c-abb}\mathbf{I}_n \end{bmatrix} = \frac{1}{c-abb}\begin{bmatrix} \frac{c-abb}{a} & \mathbf{0}_n \\ \mathbf{0}_n' & \mathbf{I}_n \end{bmatrix}$

so substituting [5-21] and [5-23] into [5-20] gives:

$$\mathbf{A}^{-1} = \left(\mathbf{L}^{-1}\right)' \mathbf{D}^{-1}\mathbf{L}^{-1}$$
$$= \frac{1}{c-abb}\begin{bmatrix} 1 & -b\mathbf{1}_n \\ 0_{\mathbf{n}} & \mathbf{I}_n \end{bmatrix}\begin{bmatrix} \frac{c-abb}{a} & \mathbf{0}_n \\ \mathbf{0}_n' & \mathbf{I}_n \end{bmatrix}\begin{bmatrix} 1 & 0_{\mathbf{n}} \\ -b\mathbf{1}_n' & \mathbf{I}_n \end{bmatrix}$$

[5-24]
$$= \frac{1}{c-abb}\begin{bmatrix} \frac{c-abb}{a} & -b\mathbf{1}_n \\ \mathbf{0}_n' & \mathbf{I}_n \end{bmatrix}\begin{bmatrix} 1 & 0_{\mathbf{n}} \\ -b\mathbf{1}_n' & \mathbf{I}_n \end{bmatrix}$$
$$= \frac{1}{c-abb}\begin{bmatrix} \frac{c-abb}{a}+nbb & -b\mathbf{1}_n \\ -b\mathbf{1}_n' & \mathbf{I}_n \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{a}+n\frac{bb}{c-abb} & -\frac{b}{c-abb}\mathbf{1}_n \\ -\frac{b}{c-abb}\mathbf{1}_n' & \frac{1}{c-abb}\mathbf{I}_n \end{bmatrix}$$

Substituting [5-12] to [5-14] into [5-24] gives the average inverse NRM of:

$$\textbf{[5-25]}\quad \mathbf{A}^{-1} = \begin{bmatrix} \frac{1}{a_{PP}} + n\frac{(s+1)^2}{4-a_{PP}(s^2+1)} & \frac{s+1}{\frac{a_{PP}}{2}(s^2+1)-2}\mathbf{1}_n \\ \frac{s+1}{\frac{a_{PP}}{2}(s^2+1)-2}\mathbf{1}'_n & \frac{2}{2-\frac{a_{PP}}{2}(s^2+1)}\mathbf{I}_n \end{bmatrix}$$

From this matrix we can derive simple rules for the construction of the inverse NRM for an open pollinated family with a degree of parental inbreeding (*F)*, where a_{pp} = 1+*F*, and a possibility (*s*) of the offspring being the result of selfing (Table 5-1). These rules can be shown to be equivalent to Henderson's (1976) rules when there is inbreeding in the parents but no selfing (*s*=0), and when there is also no inbreeding in the parents (a_{pp}=1).

Table 5-1 Rules for creation of the inverse NRM under parental inbreeding and selfing and comparison with Henderson's (1976) rules.

Part of $\mathbf{A}^{-1}$	Selfing & Inbreeding	No selfing (*s*=0)		No selfing or inbreeding (*s*=0) (a_{pp}=1, F=0)	
	These Rules	These Rules	Henderson's Rules*	These Rules	Henderson's Rules
PP	add $\frac{1}{a_{PP}}$	add $\frac{1}{a_{PP}}$	add $1/(\mathbf{L}_{PP})^2 = \frac{1}{a_{PP}}$	add 1/1=1	add 1
	add $\frac{(s+1)^2}{4-a_{PP}(s^2+1)}$ for each offspring	add $\frac{1}{4-a_{PP}}$ for each offspring	add $¼/(\mathbf{L}_{OO})^2 = \frac{1}{4-a_{PP}}$ for each offspring	add 1/(4-1)=1/3 for each offspring	add 1/3 for each offspring
OO	add $\frac{2}{2-\frac{a_{PP}}{2}(s^2+1)}$	add $\frac{2}{2-\frac{a_{PP}}{2}}$	add $1/(\mathbf{L}_{OO})^2 = \frac{1}{1-\frac{a_{pp}}{4}} = \frac{2}{2-\frac{a_{PP}}{2}}$	add 1/(1-¼)=4/3	add 4/3
PO and OP	add $\frac{s+1}{\frac{a_{PP}}{2}(s^2+1)-2}$	add $\frac{1}{\frac{a_{PP}}{2}-2}$	add $-½/(\mathbf{L}_{OO})^2$ $\frac{-1}{2\left(1-\frac{a_{PP}}{4}\right)} = \frac{1}{\frac{a_{PP}}{2}-2}$	add 1/(½-2)=–2/3	add –2/3

* the diagonal element of Henderson's **L** matrix for a parent ($\mathbf{L}_{PP}$) with inbreeding (*F*) (where a_{pp} = 1+*F*) is $\sqrt{a_{pp}}$, and the diagonal elements for the offspring ($\mathbf{L}_{OO}$) are $\sqrt{1-\frac{a_{pp}}{4}}$.

5.3.2 Testing of breeding value predictions

Partial selfing measured by outcrossing rate is commonly measured in open pollinated families (Gaiotto 1997; Patterson *et al.* 2001; Butcher and Williams 2002). We examined the effect of correcting the **A** matrix for partial selfing using Monte-Carlo simulation. Three hundred open-pollinated families were generated, each of 30 trees. Where the selfing rate was greater than zero, offspring in the family were randomly assigned to be truly open pollinated (each crossed with a different unrelated male), or to be selfs, until the desired rate of selfing was achieved. No inbreeding depression was assumed. Selfing rates of between zero and 40% were used, which corresponds to the range often found in eucalypt populations. The phenotypic variance was one, the heritabilities used were 0.2 and 0.5, and the additive and error variances were set accordingly, with no experimental design features assumed. Parental breeding values were sampled from a normally distributed population with the additive variance and a mean of zero. Offspring breeding values were calculated as the sum of the average of the female and male (self or random) parental breeding values and a random sample from a normal distributed population with a mean of zero and a variance which was half of the additive variance, the latter representing Mendelian sampling. Phenotypic observations were generated by adding to each offspring breeding value an error which was drawn from a normally distributed population with a mean of zero and the error variance. From the population, the best 180 trees were selected from both parents and offspring, with a limit of 5 trees per family (including the parents).

For the estimation of variance components, two models were run: NONE and KNOWN. For the NONE model, no selfing was assumed in the calculation of $\mathbf{A}^{-1}$. For the KNOWN model, the $\mathbf{A}^{-1}$ matrix was calculated using the actual rate of selfing. We also examined the effect of correcting the **A** matrix on the bias and accuracy of breeding values and on the efficiency of selection using another model (GIVEN), where the correct variance components were used but the $\mathbf{A}^{-1}$ was constructed assuming no selfing. This is equivalent to the variance components and pedigree used in the model applied by Jarvis *et al.* (1995). The accuracy of the breeding values for both parents and offspring was calculated as the correlation between the true and predicted values. The slope of the predicted breeding values on the true breeding values was calculated to examine the bias of the breeding values. The efficiency of

selection was measured as the gain in the true breeding values of the selections relative to the KNOWN model.

Variance components were estimated and breeding values predicted using the ASReml software (Gilmour *et al.* 1997b). A FORTRAN program (OPAINV) uses the rules derived here to form the correct inverse of the relationship matrix under parental inbreeding and partial selfing for use by ASReml. It is available from the first author on request. One thousand iterations were run for each scenario (model by selfing rate) and the average and standard error were calculated for each summary statistic.

5.4 Results

The estimates of the variance components showed that accounting for selfing in the **A** matrix with the KNOWN model leads to unbiased estimates of both the additive (Figure 5-4) and error variances (Figure 5-5), and the heritability (Figure 5-6). Failure to account for selfing with the NONE model leads to inflation of the additive variance, and deflation of the error, resulting in a net inflation of the heritability by about 20% per 10% of selfing. This confirms that for open-pollinated progeny trials with partial selfing, correcting the numerator relationship for the average rate of selfing yields correct variance component estimates.

The correlation between true and estimated breeding values shows that there is no difference between any of the models for the parental breeding values (Figure 5-7). The correlation increases with the selfing rate as more of the female parental breeding value is being expressed in the offspring. For the offspring there is no difference in the correlation between the KNOWN and GIVEN models, but the correlation for the NONE model is poorer and gets worse with increased selfing (1.0 - 1.5% per 10% of selfing) (Figure 5-8).

The slope of the relationship between the true and predicted breeding values shows that the NONE model over-predicts offspring breeding values, and the GIVEN model slightly under-predicts, when compared with the KNOWN model (Figure 5-10). The over-prediction is larger at the higher heritability (about 10% per 10% of selfing vs. 7%) and the under-prediction is smaller (0.5% per 10% of selfing vs. 2.5%). For the parental breeding values, both the NONE and GIVEN models over-predict, with the

over-prediction being greater for the NONE model (Figure 5-9). The over-prediction for the NONE model increases by about 10% per 10% of selfing, however for the GIVEN model it is higher at the higher heritability (about 6% per 10% selfing vs. 3%).

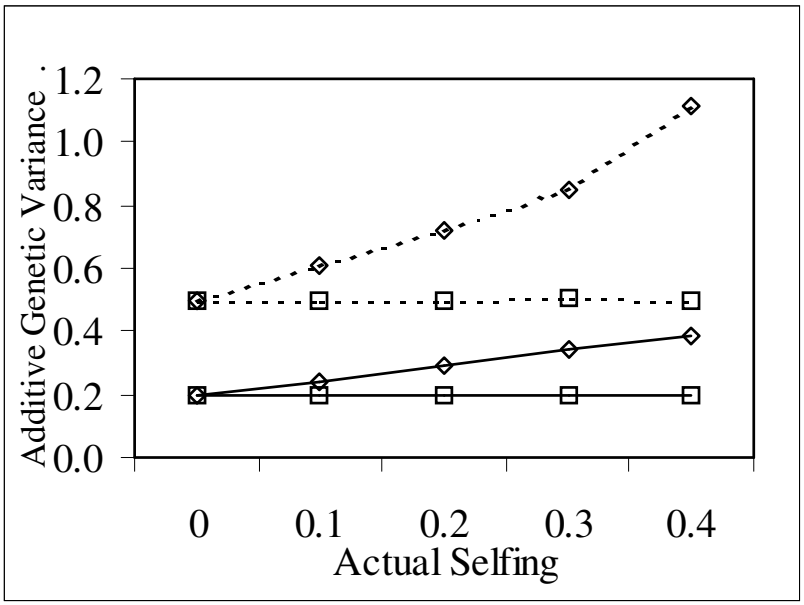


Figure 5-4 Additive variance estimates at different rates of selfing.

Numerator Relationship Matrices: NONE (◊) no selfing, and KNOWN (□) where the selfing rate is known. Heritabilities: 0.2 (——) and 0.5 (-----).

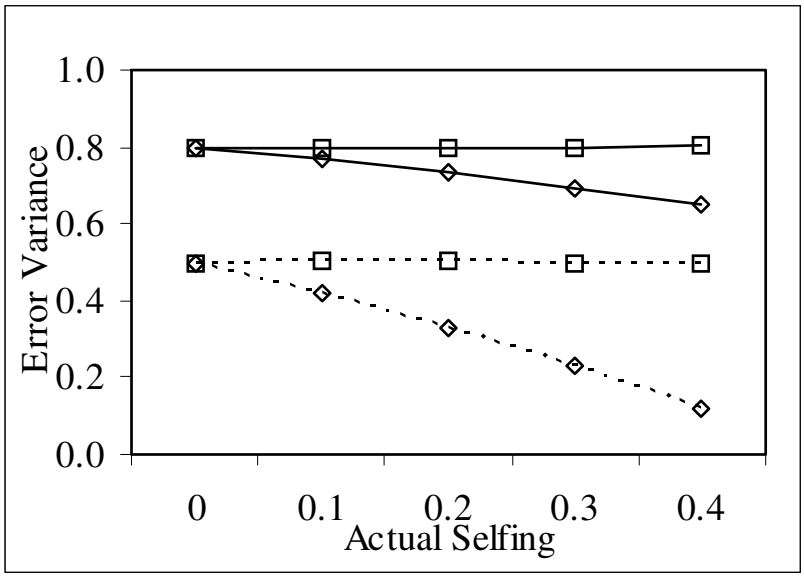


Figure 5-5 Error variance estimates at different rates of selfing.

Numerator Relationship Matrices: NONE (◊) no selfing, and KNOWN (□) where the selfing rate is known. Heritabilities: 0.2 (——) and 0.5 (-----).

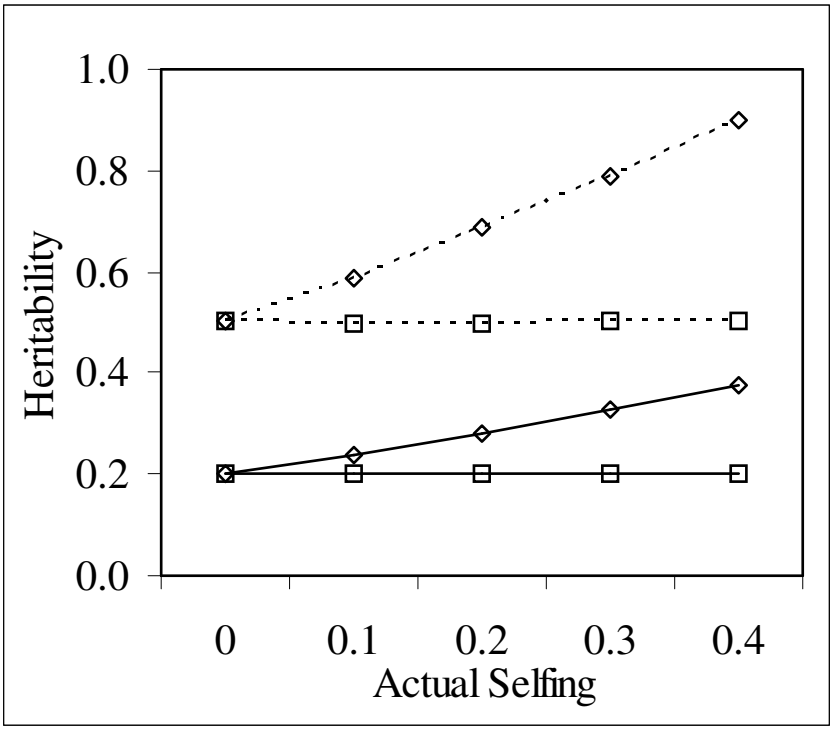


Figure 5-6 Heritability estimates at different rates of selfing.

Numerator Relationship Matrices: NONE (◊) no selfing, and KNOWN (□) where the selfing rate is known. Heritabilities: 0.2 (——) and 0.5 (-----).

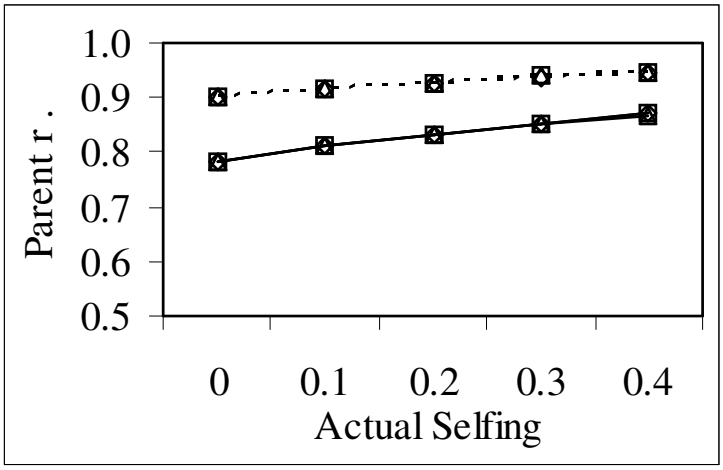


Figure 5-7 Correlation (r) between true and predicted parental breeding values for different rates of selfing.

Models: NONE (◊) no selfing, GIVEN (Δ) no selfing but correct variance components, and KNOWN (□) where the selfing rate is known. Heritabilities: 0.2 (——) and 0.5 (-----).

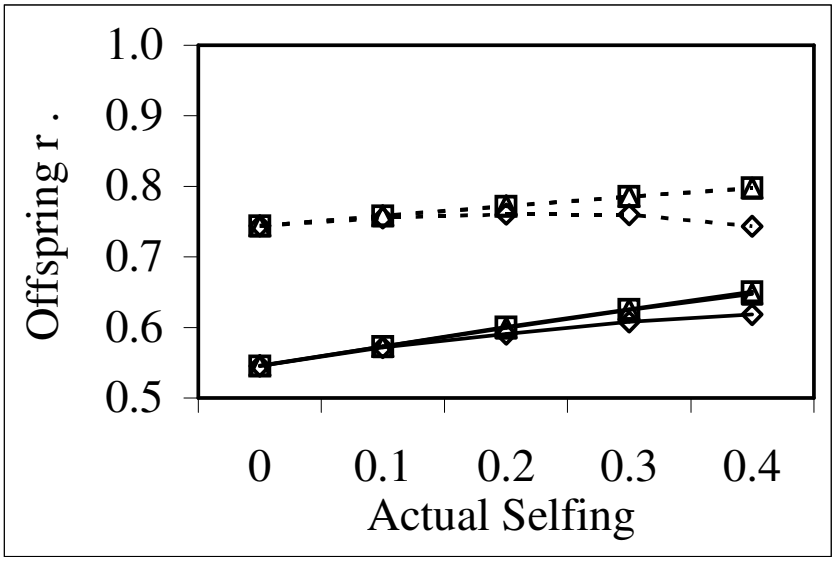


Figure 5-8 Correlation (r) between true and predicted offspring breeding values for different rates of selfing.

Models: NONE (◊) no selfing, GIVEN (Δ) no selfing but correct variance components, and KNOWN (□) where the selfing rate is known. Heritabilities: 0.2 (——) and 0.5 (-----).

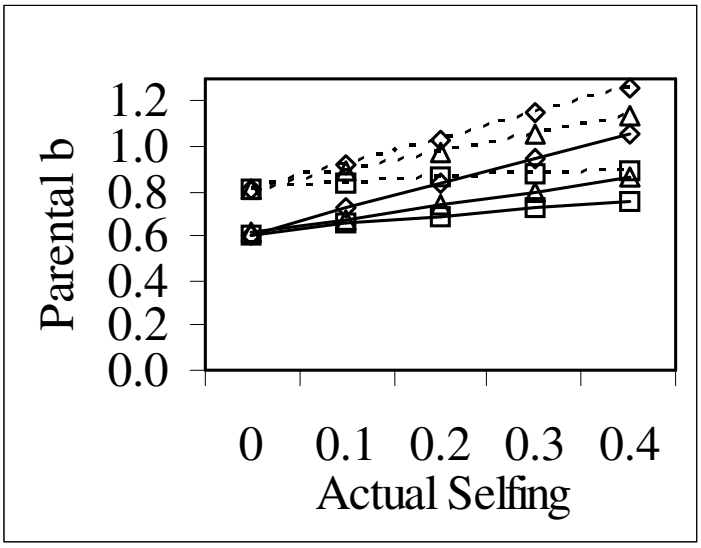


Figure 5-9 Slope (b) of the relationship between true and predicted parental breeding values for different rates of selfing.

Models: NONE (◊) no selfing, GIVEN (Δ) no selfing but correct variance components, and KNOWN (□) where the selfing rate is known. Heritabilities: 0.2 (——) and 0.5 (-----).

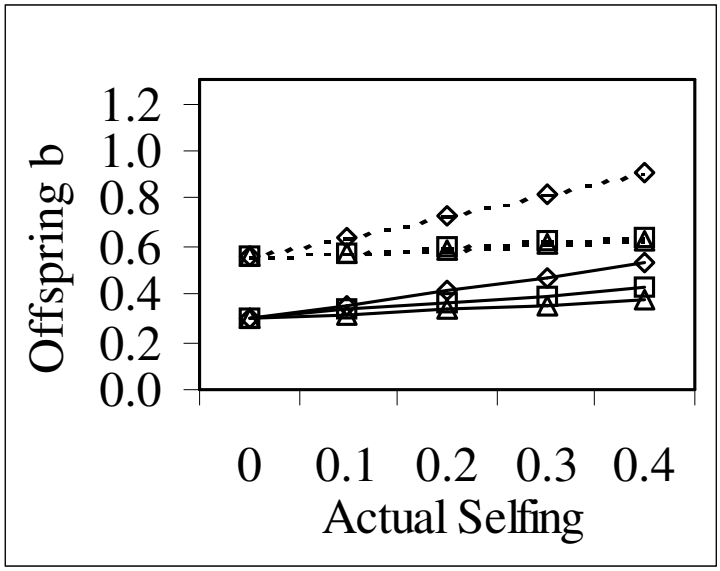


Figure 5-10 Slope (b) of the relationship between true and predicted offspring breeding values for different rates of selfing.

Models: NONE (◊) no selfing, GIVEN (Δ) no selfing but correct variance components, and KNOWN (□) where the selfing rate is known. Heritabilities: 0.2 (——) and 0.5 (-----).

The gain from selection relative to the KNOWN model indicates that there is a loss of gain with the NONE model, which increases from about 0.4% per 10% selfing at a heritability of 0.2 to about 1.6% per 10% of selfing at a heritability of 0.5 (Figure 5-11). The loss of gain for the GIVEN model is much smaller, only up to about 0.25% per 10% of selfing for the higher heritability.

5.5 Discussion

These results clearly show that it is possible to correct the **A** matrix for partial selfing in open pollinated families to give unbiased estimates of variance components. In terms of parental selection, then the model used is unimportant. The breeding values are however substantially positively biased if the incorrect relationship matrix is used, and even more so if the incorrect variance components are used. For the selection of offspring, the right variance components are clearly more important than the right relationship matrix, but both are necessary for the unbiased prediction of gain. In terms of overall gain, the losses are only modest if the incorrect relationship matrix is used, although they can be substantial if the incorrect variance components are used.

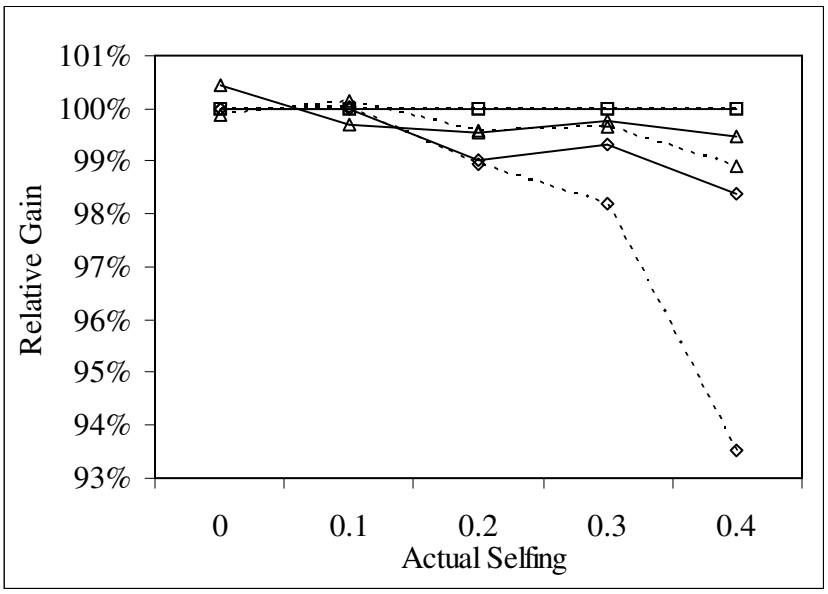


Figure 5-11 Gain relative to the KNOWN model for different rates of selfing.

Models: NONE (◊) no selfing, GIVEN (Δ) no selfing but correct variance components, and KNOWN (□) where the selfing rate is known. Heritabilities: 0.2 (——) and 0.5 (-----).

For routine analysis of breeding information correcting the relationship matrix for selfing can be easily achieved. Where the family variance is multiplied by four to provide estimates of additive and error variances then breeding values will be biased, but this will not affect selection decisions greatly. In the situation which we have simulated the gains are likely to be the smallest, as there is a very simple pedigree and only one trait. In analysis of more complex data sets where there is a mixture of open and control pollinated trees then more correct specification of the relationships will become more important for unbiased estimates of both variance components and breeding values. Additionally, in multi-trait situations, correct specification of the error covariances will rely on correctly separating the within family variation into their genetic and environmental components.

The approach that we have taken does not take into account all the problems that are inherent in the analysis of open-pollinated trials. There are a number of other factors which can increase the relationships between trees but we have only considered one of these - selfing - as a starting point. However we have seen that if the variance components are correct then although the breeding values will be positively biased selections are relatively insensitive to the structure of the relationship matrix. Any

changes brought about by different assumptions about the nature of the open-pollinated matings are likely to be even less important.

Of more importance is the likely impact of differential selfing and inbreeding depression on variance component estimation and breeding value prediction, against which the adjustments that we have made are likely to be inconsequential.

5.6 Conclusion

The rules that we have developed effectively deal with selfing for the creation of a relationship matrix, and allow unbiased prediction of variance components and breeding values.

Chapter 6 Development and comparison of methodology of spatial analysis*

6.1 Summary

Spatial analysis, using separable autoregressive processes of residuals, is increasingly used in agricultural variety yield trial analysis. Interpretation of the sample variogram has become a tool for the detection of global trend and "extraneous" variation aligned with trial rows and columns. This methodology was applied to five selected forest genetic trials using an individual tree additive genetic model. The base design model was compared with post-blocking, a first order autoregressive model of residuals (AR1), that model with an independent error term (AR1η), a combined base and autoregressive model, an autoregressive model only within replicates and an autoregressive model applied at the plot level. Post-blocking gave substantial improvements in log-likelihood over the base model, but the AR1η model was even better. The independent error term was necessary with the individual tree additive genetic model to avoid substantial positive bias in estimates of additive genetic variance in the AR1 model and blurred patterns of variation. With the combined model, the design effects were eliminated, or their significance was greatly reduced. Applying the AR1η model to individual trees was better than applying it at the plot level or applying it on a replicate-by-replicate basis. The relative improvements achieved in genetic response to selection did not exceed 6%. Examination of the spatial distribution of the residuals and the variogram of the residuals allowed the identification of the spatial patterns present. While additional significant terms could be fitted to model some of the spatial patterns and stationary variograms were attained in some instances, this resulted in only marginal increases in genetic gain. Use of a combined model is recommended to enable improved analysis of experimental data.

---

* Published as Dutkowski, G.W., Costa e Silva, J., Gilmour, A.R. and Lopez, G.A. (2002). Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research* **32**, 2201-2214.

6.2 Introduction

The usual way of dealing with site variability in both forest genetic and agricultural variety trials has been through experimental design and corresponding linear models. The traditional randomised complete block (RCB) design has largely been replaced by more sophisticated cyclic and computer-generated designs (Nguyen and Williams 1993) using incomplete blocks within replicates and models using recovery of inter-block information (Williams and Matheson 1994). However, even with sophisticated designs, there may be a mismatch between design unit boundaries and the actual patterns of site variation. Site variability in field trials can be spatially continuous, reflecting similar patterns in underlying soil and microclimatic effects; discontinuous, reflecting cultural or measurement effects; or random, because of microenvironmental heterogeneity. Spatially continuous variation may appear as a local trend (patches) or as a global trend (gradients) over the whole site.

A number of analytical approaches have been suggested to account for site variation and improve the estimation of treatment effects in the analysis of field trials. Trend surface analysis uses a polynomial function of the spatial coordinates to model environmental variation (Kirk *et al.* 1980; Tamura *et al.* 1988; Liu and Burkhart 1994). Post-blocking procedures have been used in trials with and without design structure (Ericsson 1997). Federer (1998) advocated an approach that fitted experimental design features as well as polynomials within blocks.

In this paper we focus on neighbour models that use the spatial relationship between measured units in a way analogous to time series analysis (Box and Jenkins 1970), where the data is (auto)correlated with that of its neighbours. Besag and Kempton (1986) used first differences in one spatial dimension to account for local linear trend. By assuming normality, their model could be fitted using maximum likelihood, rather than by using a simple covariate on some function of neighbour residuals (Papadakis 1937) or iteration of such a model (Bartlett 1978; Wilkinson *et al.* 1983). Gleeson and Cullis (1987) used a first-order autoregressive (AR1) function of residuals in one dimension and estimated model parameters by restricted maximum likelihood (REML) (Patterson and Thompson 1971). In an AR1 the autocorrelation ($r(X_iX_j)$) between units is a power function of their distance apart such that $r(X_iX_j) = \rho^{|i-j|}$,

where i and j are the spatial coordinates and ρ is the autocorrelation coefficient.

Cullis and Gleeson (1991) extended their method to two spatial dimensions by assuming separable AR1 processes (AR1⊗AR1, where ⊗ is the Kronecker product) in rows and columns, such that $r(X_{i,j}X_{k,l}) = \rho_{row}^{|i-k|}\rho_{col}^{|j-l|}$ for plots with row (i, j) and column (k, l) coordinates. They also used data differencing to account for global trend and advocated a model fitting procedure using spatial correlograms. Zimmerman and Harville (1991) demonstrated the close relationship between many of these approaches and proposed an approach in which global trend was modelled using fixed effects and local trend was modelled through a correlation structure based on geostatistical applications. Applying a AR1⊗AR1 model as the starting point for analysis, Gilmour *et al.* (1997a) and Cullis *et al.* (1998) suggested a model-fitting procedure where the sample variogram was used to detect global trend and "extraneous" effects (such as harvest direction) aligned with rows or columns of the trial. Once identified, appropriate model terms were added to adjust for these effects. Usually, these spatial models do not retain the experimental design features, but Williams (1986) proposed a model where only spatial relationships between plots in the same incomplete block were considered to preserve the inferential benefits of maintaining the experimental design.

Applications of spatial methods in agricultural variety yield trials have been shown to increase the accuracy (or reduce the standard error) of treatment estimates (or differences) when compared with RCB or incomplete block analyses. Neighbourhood analysis reduced the variance of variety differences by a mean of 42% (Cullis and Gleeson 1989). Clarke and Baker (1996) showed that least-squares smoothing reduced the standard error of differences by up to half, and Grondona *et al.* (1996) showed that an AR1⊗AR1 model on average halved it.

Forest genetic trials are similar to agricultural variety field trials in that there is site heterogeneity that needs to be accounted for. However, there are a number of differences. Forestry trials are often much larger than variety trials because of the large size of individual plants and the higher replication necessary to achieve satisfactory family estimates. This large size, and the fact that forestry land is often hilly and trials are broken up by site and silvicultural features, means that

environmental variability is likely to be higher in forestry tests. In forestry trials, measurements are taken on individual trees rather than plots, and tree breeders wish to select the best trees for mating, both from the parents and their offspring in the trials, rather than just make varietal comparisons. Thus, competition between trees is probably more important than inter-plot competition in variety trials. In forest genetic trials, spatial analytical methods have been used to study patterns of site variation (Fu *et al.* 1999) and have been shown to improve the precision of estimated effects for provenances (Hamann *et al.* 2001), families or parents (Magnussen 1990; Costa e Silva *et al.* 2001; Hamann *et al.* 2001), or clones (Anekonda and Libby 1996; Costa e Silva *et al.* 2001). Using simulated data, Magnussen (1993) reported that patchiness led to positively biased estimates for additive genetic variances in trials with multiple-tree plots and showed that the bias caused by spatial autocorrelation was reduced by neighbour adjustments. Competition in older trials may reduce the positive autocorrelation due to trend, especially for diameter and volume, so Magnussen (1994) has extended the Papadakis (1937) method to include competition.

In this study we contrast a number of conventional and spatial analysis approaches to forest genetic trial analysis and apply the method that Gilmour *et al.* (1997a) developed for the extended spatial modelling of agricultural variety trials. We apply the models to individual tree data, rather than plots means, with the aim of improving the prediction of breeding values for both parents and their offspring in the trials. We compare approaches in terms of model-fitting criteria, variance component estimates, the accuracy and correlation of predicted genetic values, and relative genetic gains from selection. From these analyses we develop a standard approach to spatial analysis of forest genetic trials and contrast this to agricultural variety trial analysis. The examples shown have been selected from a much larger number of trial data that we have analysed to show the application of the methods in a variety of situations we have encountered.

6.3 Materials and methods

6.3.1 Data sets

The five trials used in this study were drawn from Australia, Portugal, and Argentina and included three species (Table 6-1). Most trials were dominated by open-pollinated families, with only trial 1 having a high proportion of control-pollinated families. Most trials also contained check (usually bulk) seedlots. The trials encompassed a variety of designs, plot sizes, and replications. In trials 4 and 5, provenance was partially confounded with the family sets and, as the replicates were not contiguous, each set was treated as an incomplete block. Growth was measured as diameter at breast height (DBH) at a young age (4 or 5 years) in three trials and at age 14 years in one trial. Defoliation due to fungal infection (*Dothistroma pinii*) was scored in one trial, and stem form was assessed on a subjective four-point scale at another. For analysis, all stunted, dead, and missing trees were treated as missing values. For ease of computation, extra missing values were added to create a complete rectangular matrix of observations. The proportion of missing values ranged from 2 to 51% (Table 6-1).

6.3.2 The general statistical model

The individual tree data from each trial were all analysed using several linear mixed models of the general form:

[6-1] $$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where **y** is the vector of data, **b** is a vector of fixed effects with its design matrix **X**, **u** is a vector of random effects with its design matrix **Z**, and **e** is a vector of residuals. Fixed and random effect solutions are obtained by solving the mixed model equations (Henderson 1984):

[6-2] $$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Table 6-1 Trial and trait information.

Trial	1		2	3	4	5
Species	*Pinus radiata*		*Pinus pinaster*	*Eucalyptus globulus*	*Eucalyptus globulus*	*Eucalyptus globulus*
Country	Australia		Portugal	Australia	Argentina	Argentina
Parents			46			250
OP Seedlots[a]	20		46	123	265	250
CP Seedlots[a]	16				10	
Check lots	3		2		2	2
Grid Size[b]	19 x 84[c]		80 x 40	65 x 123	55 x 75	60 x 70
Shape	rectangle		rectangle	irregular	rectangle	irregular
Design[d]	RCB		RCB	RIB	IB	IB
Replicates	6		8	9	15	15
Rep Size	19 x 14		20 x 20	~20 x 40	non-contiguous	
Blocks				5 x 8	5 x 5	5 x 5
Plots	1 x 7		4 x 2	5 x 1	1 x 1	1 x 1
Spacing (m)	3.0 x 3.0		2.0 x 2.0	3.0 x 3.0	3.0 x 3.0	3.0 x 3.0
Trait	Defoliation	DBH	DBH	DBH	Form	DBH
Age (yrs)	2	14	5	5	3	4
Units	% as score[e]	cm	cm	mm	1-4 score	cm
N	1881	1194	3127	3948	3744	4596
Spaces (%)	19	25.2	2.3	50.6	9.2	14.5
Minimum	0.16	9.5	3.0	50	1	3.2
Mean	0.32	25.4	7.15	140.9	2.52	11.73
Maximum	0.89	42.8	13.0	300.0	4	17.5

[a] Seedlots: OP- Open pollinated, CP - Control pollinated.

[b] Size & shape: size (row x columns) and conformation for trial, replicates and plots, and tree spacing. Irregular indicates an irregular shape.

[c] every 5th row unmeasured extraction row.

[d] Design: RCB - Randomised complete block; RIB - resolvable incomplete block; IB - unresolvable incomplete block; sets are seedlot groups planted together, partially confounded with genetic groups.

[e] Subjected to arcsine-square root transformation.

where **R** is the variance–covariance matrix of the residuals and **G** is the direct sum of the variance–covariance matrices of each of the random effects. Where residuals are assumed to be independent, **R** is defined as $\sigma_e^2\mathbf{I}$, but spatial analysis allows **R** to have a different structure based on a decomposition of **e** into spatially dependent (ξ) and spatially independent (η) residuals. For models that include spatially dependent residuals we used a covariance structure that assumes separable first-order autoregressive processes in rows and columns, for which the **R** matrix is

[6-3] $$\mathbf{R} = \sigma_\xi^2\,[\mathrm{AR1}(\rho_{col}) \otimes \mathrm{AR1}(\rho_{row})] + \sigma_\eta^2\mathbf{I}$$

where σ_ξ^2 is the spatial residual variance, σ_η^2 is the independent residual variance, **I** is an identity matrix, ⊗ is the Kronecker product and AR1(ρ) represents a first-order autoregressive correlation matrix which, for ordered spatial coordinates of size *n*, has the form:

[6-4] $$AR1(\rho) = \begin{bmatrix} 1 & \rho^1 & \rho^2 & \cdots & \rho^{n-1} \\ \rho^1 & 1 & \rho^1 & \cdots & \vdots \\ \rho^2 & \rho^1 & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \cdots & \cdots & \cdots & 1 \end{bmatrix}$$

where ρ is the autocorrelation parameter. An equivalent linear model can, however, be defined where the same structure is applied to ordered row and column effects and included in **G** rather than **R**.

6.3.3 Standard models

A series of models with various combinations of random effects and definitions of **R** and **G** were evaluated for each data set to which they were applicable. The models were as follows.

1. Base, where the experimental design features of the trial were fitted. The possible models were (i) randomised complete block (RCB), where replicates and plots (where appropriate) were fitted as random effects (trials 1–3), and (ii) incomplete block (IB), where each replicate of each set was treated as a

random incomplete block effect (trials 4 and 5).

2. Resolvable incomplete block (RIB), where replicates, incomplete blocks, and plots were fitted as random effects (trial 3 only). The design was resolvable in that the trials could also be analysed as an RCB design.

3. Post-blocking (PB), where a new random incomplete block term was fitted in addition to the base (all trials except trial 3) or RIB (trial 3) models. The size of the new incomplete blocks was determined by maximization of the log-likelihood for different possible block sizes. The block sizes tried were all factorial combinations of row and column aggregations up to an arbitrary limit of 20 trees wide. Of the 400 possible combinations, only a block size of one row by one column was not fitted. The origin of the blocking system was the tree in the first row and column.

4. Autoregressive (AR1), where η was omitted, and thus all the residuals were assumed to be spatially dependent (ξ).

5. Autoregressive plus independent error (AR1η), where the residuals were assumed to be the sum of both ξ and η.

6. Combined (BaseAR1η), where the experimental design features of the base model and the autoregressive plus independent error terms were fitted together.

7. Autoregressive by replicate (AR1ηRep), which is similar to the combined model; however, the **R** matrix is structured so that there is no correlation between errors in different replicates, even though they may be spatially adjacent. This is similar to the "linear variance plus incomplete block model" model proposed by Williams (1986); however, we used an autoregressive covariance structure in R rather than a linear decay model. This model was only applied to the RCB design trials.

8. Plot autoregressive (PlotAR1η), which is similar to the AR1η model; however, the autoregressive covariance structure is applied to the plots and an independent plot error term is included. This model was applied only to trials

with multiple-tree plots.

Effects fitted to all models for a given trait were as follows.

1. A fixed genetic group effect accounted for any differences in the provenance of the parents.

2. An individual tree random additive genetic effect, with **G** including the numerator relationship matrix (Henderson 1976), modelled the genetic covariance between relatives. This allows the simultaneous estimation of breeding values for both the trees in the trials and their parents (Borralho 1995).

3. An extra independent variance for check lots of unknown parentage. Although not reported, this term was fitted to avoid bias in estimates of additive genetic variances caused by the inclusion of check trees as unrelated base trees in the numerator relationship matrix when their actual relationship was unknown. As check trees made up a substantial proportion of the data and could contribute to the estimation of environmental effects, they were retained in the data set.

4. A random set effect was also fitted in trials 4 and 5 but was found to be non-significant.

5. Missing values were fitted as fixed effects.

6.3.4 Variance parameters and model comparison

The variance parameters were estimated by REML. All non-significant ($P > 0.05$) variance parameters were eliminated from the models fitted, except for the additive genetic variance. Their significance was judged using a one-tailed likelihood ratio test (LRT) for parameters for which zero was a boundary value (Stram and Lee 1994); otherwise, a two-tailed test was used.

The models were compared using the Akaike Information Criterion (Akaike 1973):

[6-5] AIC=2*[LogL-p]

where LogL is the REML log-likelihood and p is the number of parameters estimated. Larger values of AIC reflect a better fit. Two degrees of freedom were allowed for the determination of block size in the calculation of the AIC for the post-blocking model. All design effects were fitted as random to allow for comparison using AIC, as models can only be compared when the fixed effects are the same.

6.3.5 Impact on selection

Best linear unbiased predictions (BLUPs) of parent and offspring breeding values were obtained from the solutions of the mixed model equations [6-2] using the estimated variance parameters. The accuracy of the breeding value estimates (the correlation between the true, g, and predicted, $\hat{g}$, genetic values) was calculated for each parent and offspring in the trial as:

[6-6] $$r_{g\hat{g}} = \sqrt{1 - \frac{PEV}{\hat{\sigma}_a^2}}$$

where PEV is the prediction error variance obtained from the inverse of the coefficient matrix of the mixed model equations [6-2] and $\hat{\sigma}_a^2$ is the estimated additive genetic variance.

Spearman correlations were used to compare the breeding values from the best standard model with the base model and with the extended spatial model. Relative genetic gains from selection were estimated as the difference in gain of the top 20% of parents and 5% of offspring selected by each model on the values estimated from the better model.

6.3.6 Extended spatial models

For all data sets we used the methods of Gilmour *et al.* (1997a) to derive an extended spatial model. This approach uses a perspective plot of a compressed sample variogram to detect and account for global trend and "extraneous" variation aligned with rows and (or) columns. The variogram displays the semivariance as a function of the distance between observation units. In our variograms, the residuals were separated into groups based on row and column differences to better detect effects

aligned with rows and columns. The semivariance can be computed for groups of model residuals ($\xi + \eta$) as

$$\textbf{[6-7]} \quad v_{d_{row}d_{col}} = \frac{1}{2}\frac{1}{n_{d_{row}}}\frac{1}{n_{d_{col}}}\sum_{i=1}^{n_{row}}\sum_{j=1}^{n_{col}}\left[(\xi+\eta)_{i,j} - (\xi+\eta)_{i+d_{row},j+d_{col}}\right]^2$$

where *v* is the mean semivariance for n pairs of residuals a given absolute distance d apart (called lags) in the row and column directions. If there is spatially structured variation, where observations close together are similar, then the semivariance will be low for observations that are spatially close but increases with distance until the differences become random and the semivariance plateaus. For such a situation, the variogram should rise from a value representing the independent error variance at zero lags (called the nugget) over a distance (called the range) to a plateau (called the sill) where the errors are independent. A variogram that reaches a sill is said to be stationary, an assumption that is made in the model. The software we used compressed the variogram at higher lags by aggregating lags greater than eight into sets of lags of size two, four, six, eight, and so on. The semivariance at zero lags is shown as the leading edge of a perspective plot.

To demonstrate the relationship between spatial patterns, autocorrelations, and the variogram, Monte-Carlo simulation was used to generate data from an AR1⊗AR1 process with a phenotypic variance of one on a 40 row by 40 column grid. Combinations of row and column autocorrelations between –0.9 and 0.9 were used with no independent variance, and for row and column autocorrelations between 0.3 and 0.9, the proportion of independent variance varied between zero and one. The simulations were based on a Cholesky decomposition of the variance–covariance matrix. Each realisation used the same seed for the random number generator.

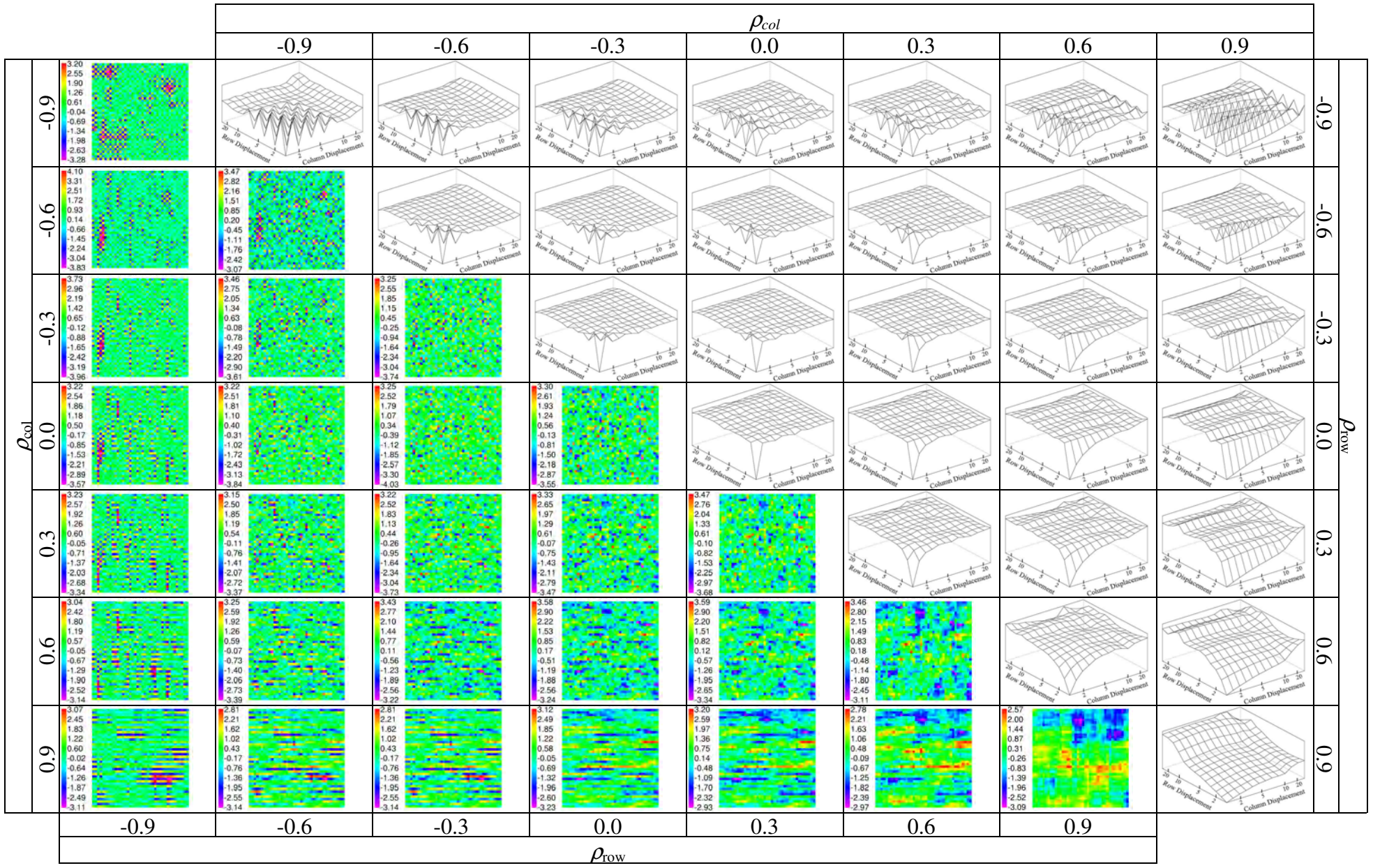


Figure 6-1 Simulated AR1⊗AR1 process for a 40 x 40 grid.

The data in plan position (lower triangle) and the sample variogram (upper triangle). The row and column autocorrelations are ρ_{row} and ρ_{col}.

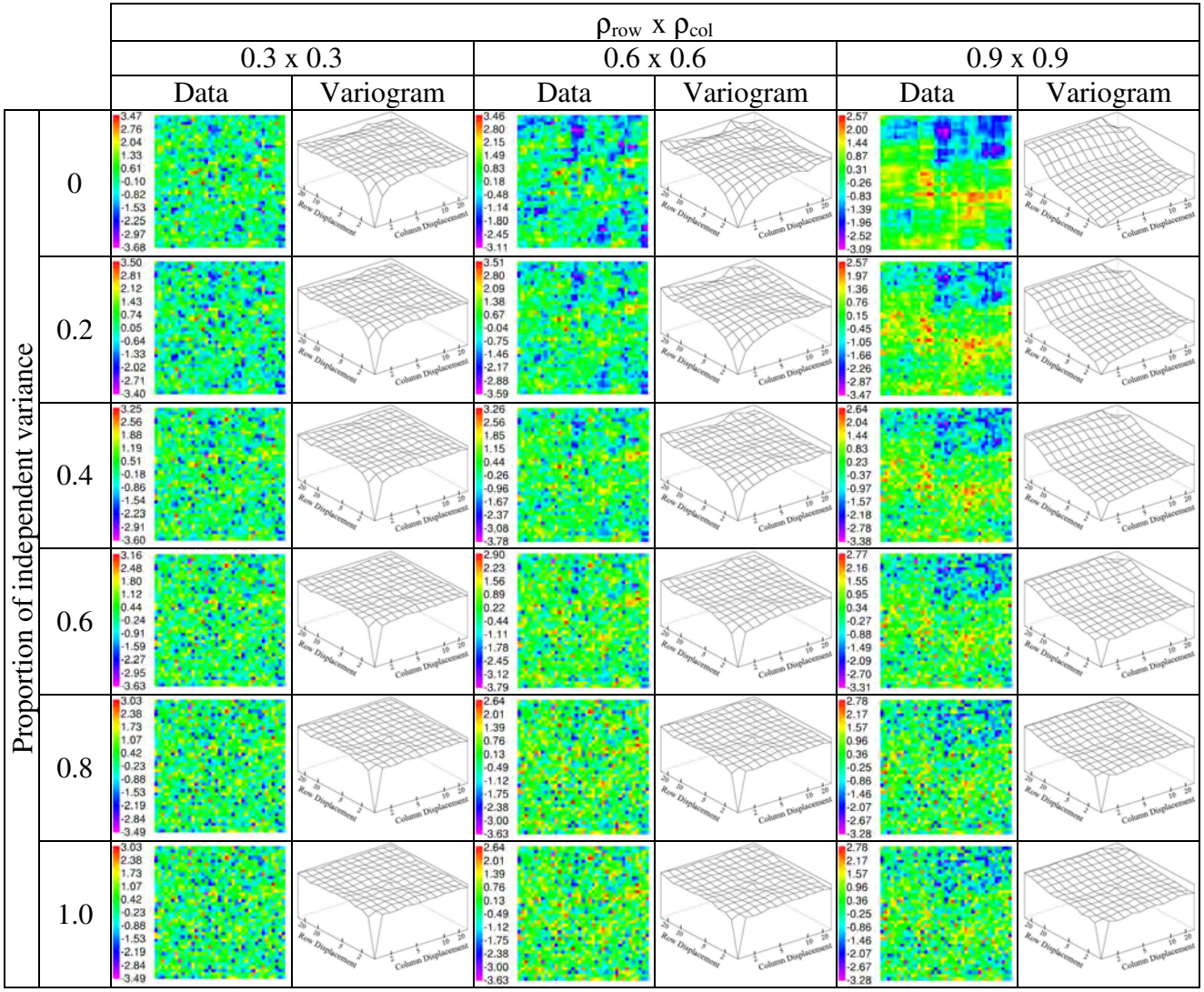


Figure 6-2 Simulated AR1⊗AR1 process with independent error for a 40 x 40 grid.
The data in plan position and the sample variogram. The row and column autocorrelations are ρ_{row} and ρ_{col}.

If there is no spatial structure ($\rho_{row} = 0$ and $\rho_{col} = 0$) then there is no pattern in the colour intensity map of the data, and the sample variogram is flat (Figure 6-1, centre). For there to be a visible pattern but with no directional structure, ρ_{col} and ρ_{row} are similar, and as they increase (down the diagonal from the centre), patches emerge at 0.3 and become more definite at 0.6. The variogram is stationary, but as ρ increases, the range increases. By 0.9 there is (in this realisation) an overall trend from top to bottom of the map and the variogram is not stationary, as the range is larger than the size of the grid used. Where there is competition, ρ_{col} and ρ_{row} are negative (up the diagonal from the centre), and an increasingly well defined check pattern becomes apparent. When ρ is negative, its even powers are positive, so adjacent values are dissimilar while those at even lags are similar. This results in a variogram with alternating high and low values at low lags, but which is flat at higher lags due to independence of the values and averaging across lags. Row or column effects occur where ρ is much stronger in one direction, and the variogram has a much longer range in that direction. For instance, when ρ_{col} is 0.9 and ρ_{row} is 0, then values in adjacent columns are similar, so a row effect becomes evident, and the variogram in the column direction does not reach the sill. The data and variogram for combinations of ρ_{col} and ρ_{row} have combinations of the characteristics from each direction. An AR1⊗AR1 process is very flexible, as it can model local and global trend, competition, and row or column effects.

If independent error is present, then the patterns in the data become increasingly unclear as the proportion of independent error increases, and the variogram develops a discontinuity at zero lags (a nugget) before becoming entirely flat with a high proportion of independent variation (Figure 6-2). Therefore, we used variograms and maps of residuals that were based only on ξ, as in all of our data sets the high proportion of η obscured any patterns.

In fitting an extended spatial model, we used the best standard model identified by the AIC as the starting point. In the variograms, global trend was detected by nonstationarity and extraneous variation was identified as ripples or ridges. Colour intensity maps of the data, residuals, and fitted effects were used to examine the different models and to help create hypotheses about other effects that could be included to improve the model. Global trend was accounted for by fitting quadratic

polynomials to the spatial coordinates or, where there was no interaction between row and column terms, by cubic smoothing splines (Verbyla *et al.* 1997). Where other effects were detected in the variogram or maps of fitted surfaces, appropriate fixed or random effects were fitted. To obtain a parsimonious polynomial model, the highest order non-significant ($P > 0.05$) polynomial terms were sequentially dropped, using an approximate incremental Wald F statistic (Kenward and Roger 1997).

All models were fitted using the ASReml software (Gilmour *et al.* 1999). Computer programs were written in FORTRAN-95 to control the ASReml runs to estimate the block size for post-blocking, for the simulation and to generate the colour intensity maps. These FORTRAN programs are available from the first author on request.

6.4 Results

6.4.1 Standard models

The Akaike Information Criterion (AIC) for the standard models (Table 6-2) showed that the RIB model was significantly better than the base model for trial 3. The postblocking model was better than the base model in every case and better than the RIB model for trial 3. The AR1 model was usually much better than the base model ($\Delta AIC > 50$), except for DBH at trial 1 where the improvement was small, and trial 4, where the AR1 model was far worse. Only for defoliation at trial 1, was the AR1 model better than postblocking. The AR1η model was better than the base and AR1 models in every case, and better than the post-blocking model for all data sets except DBH at trial 1. The combined model was the best overall, improving over the AR1η model for three data sets and reducing to the AR1η model in the other cases because of the elimination of non-significant terms. The combined model was only slightly better than the post-blocking model for DBH at trial 1. The autoregressive by replicate and the plot autoregressive models were worse than the combined model in every instance and are not considered further.

Table 6-2 Akaike Information Criterion relative to the base model.

Model	Trial – Trait (Age)					
	Trial 1 - Defol (2)	Trial 1 - DBH (14)	Trial 2 - DBH (5)	Trial 3 - DBH (5)	Trial 4 - Form (3)	Trial 5 - DBH (4)
Base	RCB	RCB	RCB	RCB	IB	IB
Block Size	10 x 19	10 x 17	15 x 13	11 x 15	15 x 5	16 x 18
RIB				76.4		
PB	71.5	12.4	119.6	118.8	28.2	154.7
AR1	86.7	5.5	71.6	58.4	-50.4	71.4
AR1η	→ 98.6	7.4	127.9	→ 147.4	48.4	→ 160.7
BaseAR1η	→ 98.6	→ 12.5	→ 133.2	→ 147.4	→ 58.5	→ 160.7
AR1ηRep	94.8	10.8	75.1			
PlotAR1η	53.0	0.0	59.3	126.8		

Base is the base model: RCB - Randomised complete block, or IB - incomplete block.

Block Size is the optimal block size in rows x columns from post-blocking.

Models: RIB - resolvable incomplete block, PB – post-blocking, AR1 – autoregressive, AR1η – autoregressive plus independent error, BaseAR1η - combined, AR1ηRep - autoregressive by replicate, PlotAR1η - plot autoregressive.

The best model for each trait is indicated with an arrow.

The best models generally showed a smooth pattern to the fitted surface, which reflected the patterns seen in the data (Figure 6-3). The exceptions were DBH at trial 1, where a checkerboard pattern dominated the surface, and for form at trial 4, where groups of vertical stripes were apparent. The boundaries of the blocks derived from post-blocking shown on the fitted surfaces generally fitted the patterns better than did those of the design features, accounting for the better performance of the post-blocking model.

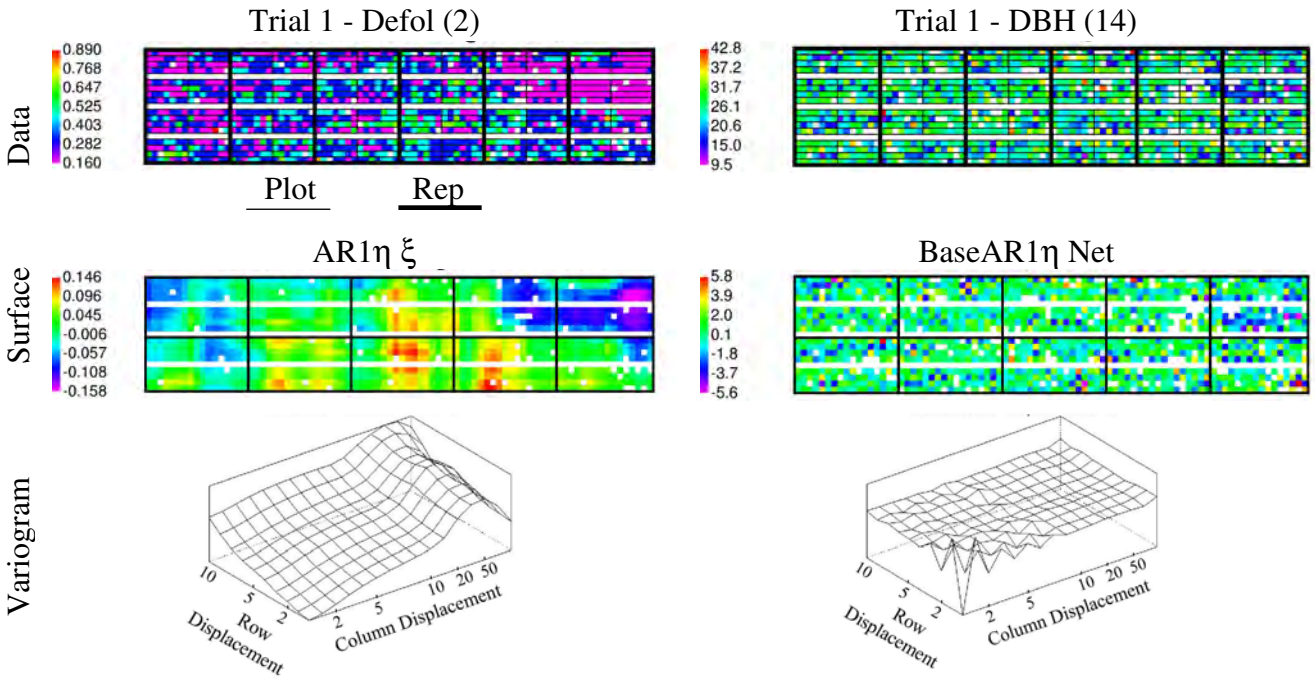


Figure 6-3 Data, best base model surface and variogram.

The data shows the boundaries of design features - plots, incomplete blocks (Inblk) and replicates (Rep). The surface is from the best base model: spatial errors from model AR1η - autoregressive plus independent errors, or the sum of spatial errors and significant design effects from model BaseAR1η - design effects plus autoregressive plus independent errors. The surface plots shows the boundaries of the blocks derived from post-blocking. The variogram is for the spatial residuals only. For Trial 1 every fifth row is an unmeasured extraction row.

Continued on next page.

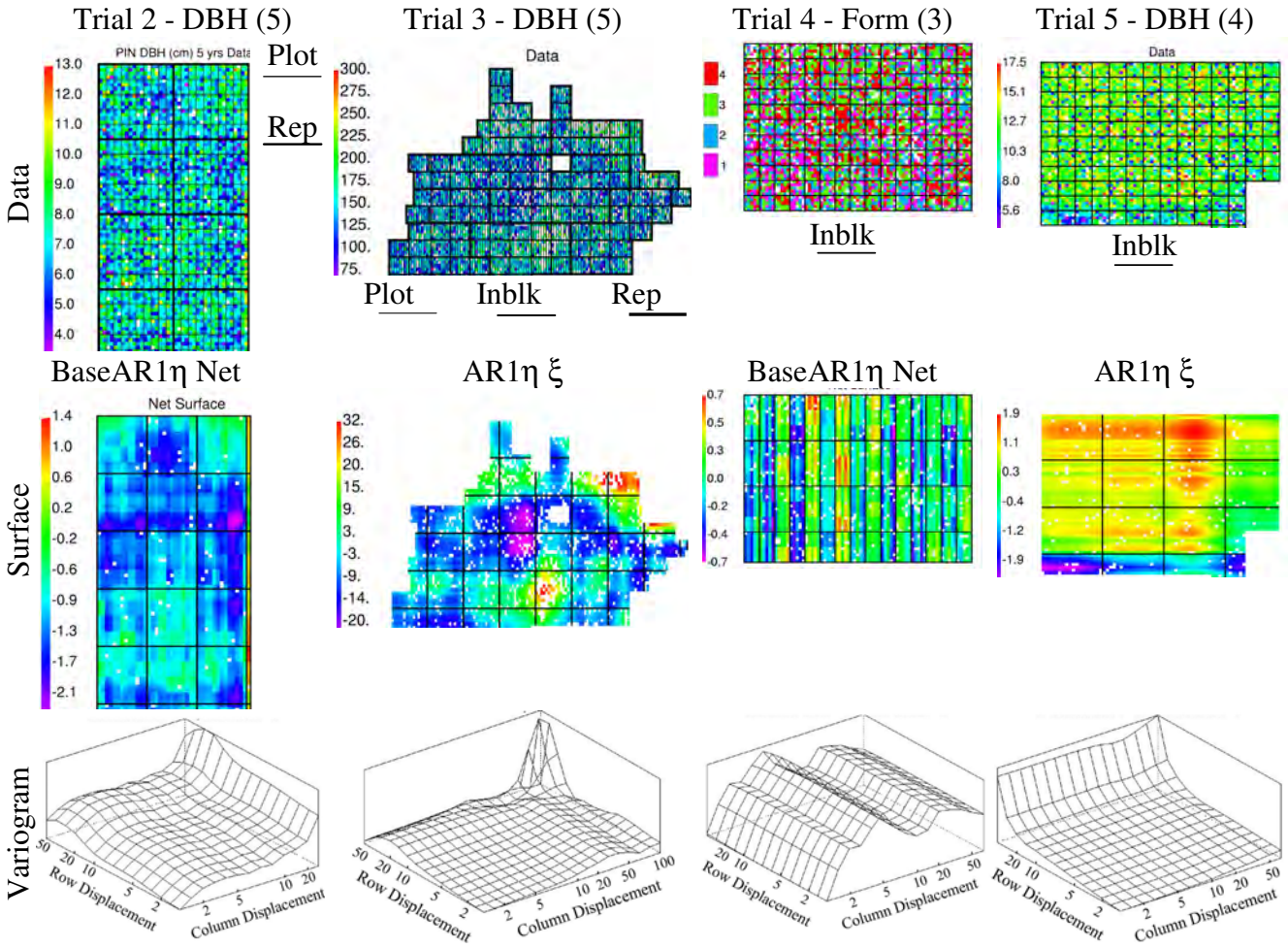


Figure 6-3 continued.

6.4.2 Extended spatial models

The variogram and map of residuals for the AR1η model for defoliation at trial 1 indicated a global trend, generally rising in both directions and with a hump in the column direction (Figure 6-3). Extending the spatial model to account for this global trend, by fitting a linear term across rows and a spline across the columns, led to a largely stationary variogram (Figure 6-4). However, the new net surface (fitted global trend plus autoregressive residuals) (Figure 6-4) showed little change from the AR1η model surface (Figure 6-3).

For DBH at trial 1 the variogram for the BaseAR1η model showed alternating ridges in both directions (Figure 6-3) as the spatial terms were fitting inter-tree competition. An extended spatial model was not pursued as competition probably operates at the

phenotypic level, so other models are more appropriate (e.g., Besag and Kempton (1986); Magnussen (1989; 1994)).

The surface for DBH at trial 2 showed markedly better growth in the last column, which coincides with an exposed edge of the trial, and this is reflected in the variogram as a raised edge (Figure 6-3). Adding a fixed effect for each of the last two columns was highly significant (F = 67.4), and led to the subsequent fitting of a spline in the row direction and an edge effect along the bottom of the trial (F = 9.71). This resulted in a stationary variogram and a fitted surface much more clearly dominated by these edge effects and a secondary trough of poor performance across the trial (Figure 6-4).

For DBH at trial 3 the variogram is dominated by a spike in the far corner (Figure 6-3) resulting from an area of fast growth. Attempts to model this trend with polynomials and splines gave no significant terms and failed to produce a more stationary variogram.

For form at trial 4 the clear column effects led to a variogram with a ridge at five lags and a trough at 10 lags (Figure 6-3), which suggests alternating groups of five columns. These groups coincide with the incomplete block boundaries and represent a single page of the assessment sheets. The assessment was, however, done by a single person assessing each individual column in a serpentine fashion, which does not explain the alternating groups of five. Fitting this column grouping as a random effect significantly improved the model (ΔAIC = 18), and the variogram now indicated stationarity (Figure 6-4). The surface now showed a combination of this column group effect and a patchy pattern (Figure 6-4).

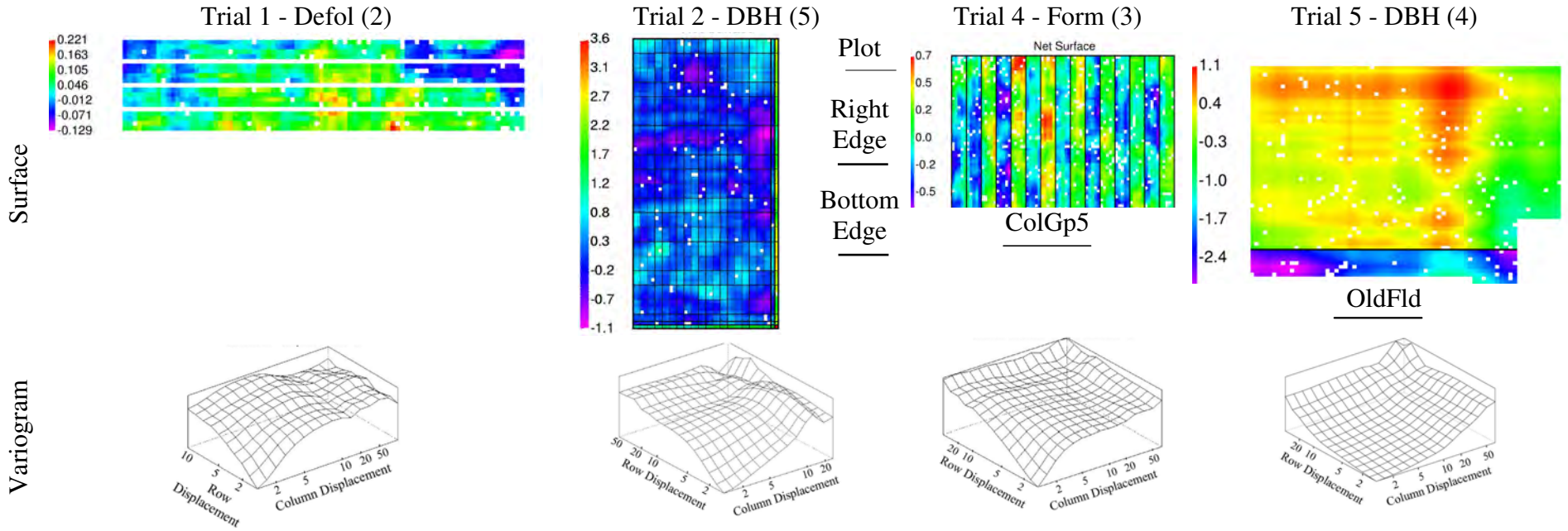


Figure 6-4 Extended model surfaces and variograms.

The surface shows the boundaries of fitted features. Trial 1 - Defol (2), AR1η + linear across rows and spline across columns (no feature boundaries are shown). Trial 2 - DBH (5), BaseAR1η (plots) + Right Edge + Spline across rows (no boundary shown) + Bottom Edge, where Right Edge are fixed effects for the last 2 columns, and Bottom Edge are fixed effects for the last 2 rows. Trial 4 - Form (3), BaseAR1η + ColGp5, where ColGP5 is a random effect for groups of 5 columns which coincide with single pages of assessment sheets. Trial 5 - DBH (4), AR1η+OldFld where OldFld is a fixed effect for an area planted into an old field. The variograms are for the spatial residuals only. For Trial 1 every fifth row is an unmeasured extraction row.

The AR1η model for DBH at trial 5 showed a zone of markedly poorer growth along the bottom of the trial, which was evident in the data and was reflected as a raised edge in the variogram (Figure 6-3). As this boundary was identified as an old fence line, beyond which more vigorous weeds in an old field led to the poorer growth, we fitted this as a fixed effect. It was highly significant (F = 49.2); however, stationarity was not achieved nor was the surface (Figure 6-4) noticeably different from the AR1η model (Figure 6-3). Attempts to model the global trend with polynomials or splines failed to produce significant terms or a stationary variogram.

6.4.3 Variance parameters and accuracy of breeding value predictions

The additive genetic variance ($\hat{\sigma}_a^2$) was significant ($P < 0.05$) with all models for all trials, although it was only marginally significant in trial 2 (Table 6-3). The additive genetic variance showed no consistent trend from the base model to the PB, best standard, and extended spatial models. In four instances $\hat{\sigma}_a^2$ increased markedly with the AR1 model. This increase, however, was clearly an artefact of an inappropriate model, as $\hat{\sigma}_a^2$ decreased again with the better AR1η model. Random tree-to-tree variation was inflating the additive genetic variance as there was no other term that could effectively account for it.

The design effects were significant ($P < 0.05$) in the base models except for DBH in trial 1. Plot effects decreased with the RIB model, and all design effects decreased markedly or became non-significant with the PB model. For the BaseAR1η model, all design effects were eliminated in three cases, and only plot effects remained in two others, with a much reduced plot variance in trial 2.

Table 6-3 Variance parameter estimates and breeding value accuracies for selected standard models and extended spatial models.

All variances for each data set are scaled by the error variance for the base model. The variances ($\hat{\sigma}^2$) are for replicates (r), incomplete blocks (i), blocks from post-blocking (and column groups for extended spatial model in trial 4) (b), plots (p), additive genetic effects (a), and errors: from non-spatial model (e), and from spatial models, independent (η), and autoregressive errors (ξ). ρ_{row} and ρ_{col} are the row and column autocorrelations. The superscripts are an approximate t statistic (variance/standard error). $r_{g\hat{g}}$ is the breeding value accuracy for parents and trees in the trial. Models: Base - base model (RCB - Randomised complete block, or IB - incomplete block), RIB - resolvable incomplete block, PB – post-blocking, AR1 - autoregressive, AR1η - autoregressive plus independent error, BaseAR1η - combined, Best - best of the standard models, and Extended - extended spatial model (where fitted).

Trial-Trait (Age)	$\hat{\sigma}_r^2$	$\hat{\sigma}_i^2$	$\hat{\sigma}_b^2$	$\hat{\sigma}_p^2$	σ_e^2	σ_η^2	σ_ξ^2	ρ_{row}	ρ_{col}	σ_a^2	$r_{g\hat{g}}$	
Model											Par-ents	Trees
Trial 1-Defol (2)												
Base - RCB	0.170[1]			0.286[5]	1.000[5]					1.036[3]	0.708	0.723
PB	ns		0.442[2]	0.112[3]	0.979[5]					1.015[4]	0.737	0.739
AR1							0.674[5]	0.835[19]	0.849[22]	2.163[20]	0.781	0.943
AR1η (Best)						0.884[5]	0.654[4]	0.875[21]	0.877[24]	0.912[4]	0.735	0.727
BaseAR1η	ns			ns		0.884[5]	0.654[4]	0.875[21]	0.877[24]	0.912[4]	0.735	0.727
Extended						0.770[4]	0.499[5]	0.658[8]	0.715[9]	0.908[4]	0.731	0.727
Trial 1-DBH (14)												
Base - RCB	ns			0.044[3]	1.000[11]					0.266[3]	0.660	0.581
PB	ns		0.042[2]	ns	0.964[10]					0.320[3]	0.697	0.619
AR1							1.040[11]	-0.13[3]	ns	0.267[3]	0.680	0.593
AR1η						0.922[10]	0.109[2]	-0.47[2]	-0.80[7]	0.273[3]	0.687	0.598
BaseAR1η (Best)	ns			0.072[3]		0.682[4]	0.315[2]	-0.40[2]	-0.37[2]	0.238[2]	0.647	0.568

Continued on next page.

Table 6-3 Continued.

Trial-Trait (Age) Model	$\hat{\sigma}_r^2$	$\hat{\sigma}_i^2$	$\hat{\sigma}_b^2$	$\hat{\sigma}_p^2$	σ_e^2	σ_η^2	σ_ξ^2	ρ_{row}	ρ_{col}	σ_a^2	$r_{g\hat{g}}$ Par-ents	Trees
Trial 2-DBH (5)												
Base - RCB	0.018[1]			0.124[6]	1.000[23]					0.065 [2]	0.567	0.346
PB	ns		0.396[3]	0.071[5]	0.976[25]					0.053 [2]	0.576	0.344
AR1							0.698[3]	0.979[136]	0.941[47]	1.284[33]	0.959	0.977
AR1η						0.959[25]	0.864[3]	0.984[166]	0.955[58]	0.077 [2]	0.693	0.418
BaseAR1η (Best)	ns			0.041[3]		0.959[26]	1.651[2]	0.993[369]	0.978[89]	0.053 [2]	0.587	0.351
Extended	ns			0.029[2]		0.916[22]	0.106[4]	0.735 [9]	0.831[12]	0.066[2]	0.633	0.380
Trial 3-DBH (5)												
Base - RCB	0.080[2]			0.116[6]	1.000[23]					0.199 [5]	0.763	0.468
RIB	0.081[1]	0.092[5]		0.032[2]	0.996[24]					0.206 [5]	0.786	0.493
PB	0.034[1]	0.041[3]	0.088[3]	0.031[2]	0.980[23]					0.208 [5]	0.790	0.498
AR1							0.280[4]	0.922[37]	0.945[37]	1.488[41]	0.926	0.971
AR1η (Best)						0.963[23]	0.273[4]	0.923[37]	0.947[37]	0.213 [5]	0.801	0.510
BaseAR1η	ns			ns		0.963[23]	0.273[4]	0.923[37]	0.947[37]	0.213 [5]	0.801	0.510
Trial 4-Form (3)												
Base - IB		0.090[6]			1.000[29]					0.086[4]	0.544	0.224
PB		0.020[2]	0.072[4]		1.000[29]					0.086[4]	0.547	0.229
AR1							1.093[30]	0.115 [6]	0.103[6]	0.081[3]	0.531	0.199
AR1η						0.978[28]	0.132 [4]	0.988[158]	0.572[4]	0.084[3]	0.542	0.222
BaseAR1η (Best)		0.030[3]				0.966[28]	0.086 [4]	0.995[194]	0.381[2]	0.084[4]	0.543	0.224
Extended		ns				0.938[24]	0.092[4]	0.750 [8]	0.581[5]	0.083[4]	0.545	0.233
Trial 5-DBH (4)												
Base - IB		0.161[6]			1.000[17]					0.475 [7]	0.797	0.603
PB		ns	0.277[3]		0.985[17]					0.483 [7]	0.803	0.614
AR1							0.507[2]	0.963[50]	0.994[307]	1.723[41]	0.916	0.984
AR1η (Best)						0.958[16]	0.582[2]	0.970[57]	0.995[294]	0.485 [8]	0.805	0.617
BaseAR1η		ns				0.958[16]	0.582[2]	0.970[57]	0.995[294]	0.485 [8]	0.805	0.617
Extended						0.962[16]	0.120[2]	0.960[39]	0.971 [40]	0.485 [7]	0.805	0.614

The error variance ($\hat{\sigma}_e^2$) decreased slightly for the IB and PB models, except for form at trial 4, where it was unchanged. With the AR1 model the autocorrelations ($\hat{\rho}$) were similar in the row and column directions and, where the $\hat{\rho}$'s were close to one (i.e., all cases except DBH at trial 1 and form at trial 4), $\hat{\sigma}_\xi^2$ decreased markedly compared with the error variance of the base model. With the AR1η model most of the error variance present was spatially independent as the ratio $\hat{\sigma}_\eta^2 / \hat{\sigma}_\xi^2$ ranged from 1.1 to 7.4 (not shown). In four instances $\hat{\sigma}_\xi^2$ and $\hat{\rho}$ only changed slightly between the AR1 and AR1η models. However, for DBH at trial 1 and form at trial 4, $\hat{\sigma}_\xi^2$ was reduced dramatically and the $\hat{\rho}$'s became more negative or positive, as well as more asymmetric. These two trials had the highest $\hat{\sigma}_\eta^2 / \hat{\sigma}_\xi^2$ ratios, and omitting the independent error in the AR1 model obscured the actual patterns of variation, as indicated by low AIC values (Table 6-2), and lead to large changes in the estimated spatial parameters (Table 6-3). The residual variances and the $\hat{\rho}$'s decreased between the best standard and extended spatial models, except for form at trial 4 where $\hat{\rho}_{\text{row}}$ went down and $\hat{\rho}_{\text{col}}$ increased to become more similar as the strong column group effect was accounted for.

Except for DBH at trial 1, the accuracy of both parent and offspring breeding values increased between the base and best standard models, with a mean increase of 2.9% for parents and 2.7% for offspring (Table 6-3). Where the additive genetic variance was inflated with the inappropriate AR1 model, the accuracies were also inflated. There was no consistent change in accuracies between the best standard model and the extended spatial model.

6.4.4 Breeding value correlations and relative genetic gains

There were small but consistent differences in predicted breeding values between the base and best standard models (Table 6-4). As expected, the correlations were all higher than 0.95 (except in trial 2), and the relative genetic gains were mostly less than 6%. There was almost no gain from the extended spatial model above the best standard model: parental gains were usually zero and the largest offspring gain was around 1%.

Table 6-4 Correlation of breeding values and gains in selection between models.

The models are: Base – base model, Best - best standard model, and Extended - extended spatial model (where fitted).

Trial - Trait (Age)	Model 1	Model 2	Breeding value correlation		Selection gain	
			Parents	Trees	Parents 1 in 5	Trees 1 in 20
Trial 1 - Defol (2)	Base	Best	0.970	0.968	5.69%	1.67%
	Best	Extended	0.997	0.998	0.00%	0.27%
DBH (14)	Base	Best	0.998	0.994	0.00%	0.94%
Trial 2 - DBH (5)	Base	Best	0.922	0.936	2.77%	5.06%
	Best	Extended	0.986	0.987	0.00%	0.91%
Trial 3 - DBH (5)	Base	Best	0.965	0.966	3.42%	4.36%
Trial 4- Form (3)	Base	Best	0.989	0.988	0.04%	1.72%
	Best	Extended	0.994	0.993	0.43%	0.98%
Trial 5 - DBH (4)	Base	Best	0.991	0.988	1.11%	1.57%
	Best	Extended	0.999	0.999	0.14%	0.15%

6.5 Discussion

On the basis of these results, we advocate an initial combined model for spatial analysis of forest genetic trials which adds an autoregressive error term to the design model and retains an independent error term. In most instances, this was a considerably better model, and although the differences in selected trees were not large, the model does change the trees that are selected. Although not so different

from the alternative models that we tried, it is simple to apply and does not inflate the additive variance. The elimination of non-significant effects from the combined model may reduce the combined model to one without design effects terms, but in a few instances they are retained. While there is no a priori way of identifying which ones this might be, our results, and those of Costa e Silva *et al.* (2001), suggest that where variation is purely continuous in nature, then most design effects will be much reduced if not eliminated. However, the design effects may remain significant in the presence of autoregressive error terms when there is inter-tree competition or some other unusual effects present.

Gilmour *et al.* (1997a) suggested an initial AR1⊗AR1 model for agricultural variety plot trials. Early generation agricultural variety trials are often unreplicated and may have no design features to retain (e.g., Cullis and Gleeson (1989); Cullis *et al.* (1998); Moreau *et al.* (1999)). However, Qiao *et al.* (2000) found that in most of the 33 designed wheat trials examined design features were retained, and others have also advocated a combined approach (Federer 1998).

Both an independent and an autoregressive error are necessary. These data, and those of Kusnandar and Galwey (2000) and Costa e Silva *et al.* (2001), clearly show that in forestry trials the independent error is always present, is large, and accounting for it is always necessary. When fitting an individual tree additive genetic model, without an independent error term, the additive genetic variance can be substantially inflated and the actual patterns of spatial variation may be obscured.

In variety trials where the experimental unit is the plot, an independent error is assumed to represent measurement error. If it is modelled, it is often significant but usually small (Basford *et al.* 1996; Gilmour *et al.* 1997a; Cullis *et al.* 1998). In forestry trials, while measurement error may exist, variation from tree to tree will also be due to microsite and non-additive genetic effects, two sources of variation that are not present in the plot means of the inbred lines used in variety trials. As the large independent error persists in clonal trials (Costa e Silva *et al.* 2001) and across a range of types of measurement data, this suggests that microsite variation is its primary cause.

We advocate no routine further modelling of trends detected in the initial spatial model. While the tools we used greatly aided in the understanding of the nature of variation at a trial site, the extended spatial modelling required a great deal of effort and did not change the breeding values or selections substantially. Our lack of success compared with such efforts in variety trials (Gilmour *et al.* 1997a; Cullis *et al.* 1998; Qiao *et al.* 2000) may be due to the much larger array sizes we have used, the more heterogeneous nature of forestry trials and the high proportion of independent error present. Also, the AR1⊗AR1 model seems to be extremely robust and able to adequately accommodate a wide variety of the situations, as Figure 6-1 demonstrates. However, one cannot be dogmatic; other effects can be detected, and where they lead to estimation of significant effects and changes in breeding values, they should be considered.

Not fitting detected trends may lead to problems. Global trends and effects aligned with rows or columns will inflate autocorrelations, which may mask local trends. Brownie and Gumpertz (1997) recommended on the basis of simulation studies that, where present, global trends should be fitted. Failure to do so may lead to estimates of precision that are too small. On the other hand, there is a danger of overfitting effects and artificially reducing the estimates of precision. Although we compared a large number of models, including a great number to arrive at the best block size for postblocking, our measure of model fit (AIC) does not account for multiple comparisons. There is also a difficulty in correctly defining the number of degrees of freedom in a spatial analysis. However, the large AIC differences from adding autoregressive residuals to the design model and the consistency of the results with the data patterns gives confidence to the conclusions we have drawn.

Previous spatial analyses of forestry data have indicated that the autocorrelations were low. Magnussen (1990) found autocorrelations of between 0 and 0.4 for height plot means in Jack Pine and used values up to 0.5 in simulation studies (Magnussen 1993, 1994). Anekonda and Libby (1996) found values up to 0.4 in a variety of clonal redwood traits, and Dutkowski *et al.* (unpublished data) found values up to 0.25 in plot basal area for *Eucalyptus globulus*. However, our much higher autocorrelations are consistent with these studies because they are calculated on a different basis. Our model separates independent error from the much smaller autocorrelated

component, whereas previous studies have used a single autocorrelated error. Fitting a family model with no independent error term (not shown) led to much lower autocorrelations.

Our results indicated that the spatial model had no consistent effect on the additive genetic variance. This is broadly consistent with the results from analysis of plot data by Magnussen (1990) but not with his simulation results (Magnussen 1993, 1994), which indicated that patchiness inflated the additive genetic variance. This may again be due to the very high independent error that we have found.

Post-blocking was better than the design model in many cases, confirming the results of Fu *et al.* (1999), who found that row or column factors explained more variation in Douglas Fir progeny trials than the replicates did. Even though some of our selected cases displayed patterns aligned with rows or columns which would suit post-blocking, it was also better than the design model where continuous trend dominated. It, however, was not as good as the combined model. Where computing resources are limited and large across-site breeding value predictions are being undertaken (e.g., Jarvis *et al.* (1995)), post-blocking may be an appropriate approach. However, such computing restrictions are rapidly decreasing, and if they do apply, then for each site the fitted spatial surface may be simply subtracted from the data to greatly simplify across-site predictions. However, both of these approaches may be inefficient where replication at any site is insufficient to properly estimate the environmental surface. The problems inherent in multiple comparisons of spatial models also apply to the post-blocking procedure that we have used.

Finally, the confirmation of efficiency and validity that has been demonstrated through simulation in the agricultural variety trial situation (Lill *et al.* 1988) (Baird and Mead 1991; Cullis *et al.* 1992; Brownie and Gumpertz 1997; Azais *et al.* 1998) needs to be verified for forest genetic trials. We are currently looking at these issues through simulations with assumptions that better match the situations we have found.

6.6 Conclusion

We believe that spatial analysis has a major role to play in the analysis of forest genetic trials. While the gains presented here have been modest, other analyses (Costa e Silva *et al.* 2001) have indicated that much more substantial gains can be achieved. We advocate analysis at the individual tree level, retaining the design effects, including an independent error term, and not pursuing extended spatial models. This approach differs from that advocated for agricultural variety field trials because of their differences from forest genetic trials. Although the advantage of using spatial analysis is diminished with good design, an autoregressive error structure usually accommodates the spatial variation more naturally than designed blocks. There are many trials designed with relatively large blocks for which it would be vital. Even for well-designed trials adding spatial terms may improve the model. The power of spatial analysis to reveal hitherto undetected trends and improve understanding of variation is an added bonus. Understanding the nature of variation can only help in our design, measurement and analysis of trials.

6.7 Acknowledgements

We thank the Estação Florestal Nacional, Lisbon, Portugal; WA Plantation Resources, Manjimup, Western Australia; Hancock Victorian Plantations, Melbourne, Victoria, Australia; and Instituto Nacional de Tecnología Agropecuaria (INTA), Buenos Aires, Argentina, in association with Sociedade Portuguesa de Papel (SOPORCEL), Portugal, for making the data available. For his useful comments on the manuscript, we thank Luis Apiolaza. We also thank the Fundação para a Ciência e Tecnologia, Lisbon, Portugal, for their financial support of the second author.

Chapter 7 Application of spatial analysis methodology to many data sets*

7.1 Summary

Spatial analysis of a wide variety of forestry trials and variables has resulted in improvements of more than 10% in predicted genetic responses for around one tenth of the 216 variables tested, although in general the gains were more modest. This spatial method augments the standard analysis model by modelling the residuals as the sum of an independent component and a two dimensional spatially auto-correlated component, and is fitted using REML. The largest improvements from the augmented model were for tree height. Traits with little spatial structure to the environmental variation, such as stem counts, and form and branching scores, did not respond as often to spatial analysis. The spatially auto-correlated component represented up to 50% of the total residual variance, usually subsuming any variation explained by design based blocking effects. The auto-correlation parameter, which is related to patch size, tended to be high for growth, indicating a smooth surface, small for measures of health associated with insect and disease attack, indicating patchiness, and intermediate for other traits. Competition effects, indicated by negative spatial auto-correlations, were dominant in only 10% of diameter measurements. They were detected in square planted trials with an average diameter greater than 17cm, and between the closer trees in trials with rectangular spacing. The likelihood surface was sometimes found to be bimodal, indicating that competition may be present, but not dominant, in more cases. Modelling of extraneous effects such as assessment direction detected in trials with severely asymmetric auto-correlations yielded extra genetic gain in only the most extreme cases. More sophisticated resolvable incomplete block or row-column designs were better than randomised complete block

* Sumbitted for publication as Dutkowski, G.W., Costa e Silva, J., Gilmour, A.R., Wellendorf, H., and Aguiar, A. (2005) Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials.

designs, but adding auto-correlated spatial terms usually resulted in further improvement in model fit.

7.2 Introduction

The estimation of treatment effects corrected for environmental effects is the primary aim of field experiments. Spatial analysis using two dimensional separable auto-regressive error models, with parameters estimated using restricted maximum likelihood (REML) (Cullis and Gleeson 1991), has become commonplace in agricultural variety trial analysis (Braysher *et al.* 2001; Singh *et al.* 2003; Yang *et al.* 2004). In such a model the auto-correlation between observations is a power function of their distance apart. Extra global trend and extraneous effects (such as harvest direction) can be detected using a variogram and added to the model (Gilmour *et al.* 1997a). This, and related spatial methods such as the Papadakis neighbour residual covariate approach (Papadakis 1937), or least squares smoothing (Smith and Casler 2004; Yang *et al.* 2004), have been shown to increase the accuracy of treatment estimates (or differences), and to increase the treatment correlation between tests (Clarke and Baker 1996; Qiao *et al.* 2003).

Recent work on spatial analysis in forest genetic trials (Costa e Silva *et al.* 2001; Dutkowski *et al.* 2002), based on the work of Gilmour *et al.* (1997a), has recommended that the standard analytical model using experimental design features be augmented with a spatial component in the form of separable two-dimensional (anisotropic) auto-regressive residuals. This model gave significant improvement over standard design models in all instances tested, and in some cases gave substantial selection gains. It differed from agricultural variety trial analysis in that an individual tree additive genetic model (Henderson 1976) was used, design terms were retained, an independent error term was always necessary and usually much larger than the spatial component, and attempts at further modelling of global trend and extraneous sources of variation were not successful or yielded little or no selection gain.

While this particular approach has not yet been adopted in forest genetic trials, it remains an area of much interest. Since the work of Magnussen (1990) several methods have been reported for both tree and plot data. Kusnadar and Galwey (2000)

and Ipinza and Gutiérrez (1999) used models similar to ours, confirming the dominance of the independent error for individual tree data. Williams *et al.* (2005) defined a separable model with a linear decay of auto-correlation within replicates. They found that fitting global trend (using splines and local effects) and an auto-regressive residual model, after the method of Gilmour *et al.* (1997a), gave much the same results as their model for provenance trial plot data. Also for provenance trial plot data, Saenz-Romero *et al.* (2001) found that a mixed model incorporating trend surface analysis and an isotropic exponential auto-correlation decay model was better than the design model or trend surface analysis on its own. Joyce *et al.* (2002) applied a two-stage approach, first adjusting the residuals from a family model for global trend by estimating row and column effects using median polishing, and then using a neighbour adjustment based on patch sizes estimated from isotropic variogram modelling to account for local trend. This approach grew out of the work of Fu *et al.* (1999) where row and column effects were detected in a large proportion of trials analysed and auto-correlation was found in the residuals using an isotropic variogram. Anekonda and Libby (1996) used a similar adjustment for local trend, but without the use of the variogram to establish patch size. Hamann *et al.* (2001) modelled the residuals from a family model using a spherical isotropic variogram model, and used the estimated kriging surface to adjust the data. Mora-Garces and Ramirez (2000) have tried both trend surface analysis and the Papadakis method, and De Souza *et al.* (2003) also compared the Papadakis method with a number of other approaches.

In this study we applied the method used by Costa e Silva *et al.* (2001) and Dutkowski *et al.* (2002) to 216 variables from 55 trials in order to gain a better understanding of the utility of spatial analysis in a wide range of situations. The larger data set enabled us to see if the patterns observed in our previous small-scale studies were general. Height and diameter measurements were analysed across a range of tree sizes. A wide variety of other traits were also analysed. Several trial designs were represented, and spatial analysis was evaluated against trial designs having within-replicate blocking factors (Nguyen and Williams 1993).

7.3 Materials and methods

7.3.1 Data sets

The data sets used in this study were from 55 trials representing six species grown in five countries, but were principally from Australia and Denmark (Table 7-1). The trials were primarily chosen because the trees could be easily mapped on a rectangular grid. Other considerations were that the trials covered a range of designs, and had data on various traits from trees of various sizes. Results of spatial analyses from some of these trials have been previously reported in Costa e Silva *et al.* (2001) (PS01, PP01, PR08B, 9B, 9A, 5B and 1B), Dutkowski *et al.* (2002) (PR05A, PP01, EG10D, EG02 and EG01), and Joyce *et al.* (2002) (PM01). Trial codes identify the genus, species (e.g. PR = *Pinus radiata*, see Table 7-1) and site number, with trials on the same site have different alphabetic suffices.

The trials were predominantly randomised complete block (RCB) designs (Table 7-1). There was one completely randomised design (CR). Three trials were termed trend surface (TS) designs, where the treatment plots were completely randomised, but with over-replication of check lots. This is similar to the unreplicated variety trials analysed by Cullis *et al.* (1989), but in this case they were specifically designed to allow estimation of the underlying environmental surface using trend surface analysis (Tamura *et al.* 1988). Trial series PR06 and PR07 similarly had over-replication of check lots, but within the framework of a RCB design. They also formed a series over a range of tree spacings from 2.1x2.1 to 4.2x4.2m. There were twelve trials with resolvable incomplete block (RIB) designs (Patterson and Williams 1976), where the treatments were balanced across replicates, as well as two trials with incomplete block (IB) designs without contiguous replicates. The most sophisticated design was the resolvable row-column (RRC) design (Nguyen and Williams 1993) used in six of the trials.

Table 7-1 Summary of trials used in the analysis.

Code[1]	Country	Pedigree[2]	Des-	Size[4]	Reps		Blks	Plots[5]	Spacing	Fill	Variables[8]		
		OP/CP/Chk/P	ign[3]		n	Size[4]			(m)[6]	(%)[7]	n	Ht	Diam
EG01	Argentina	250/0/0/251	IB	60x70	15	-	150	1	3.0x3.0	12-24	8	2	3
EG02	Argentina	265/10/0/271	IB	55x75	15	-	165	1	3.0x3.0	7-38	5	1	2
EG03	Australia	53/0/0/53	RCB	18x48	12	9x8		1	2.6x3.5	2-5	3	2	
EG04	Portugal	57 clones	RCB	8x37	4	2x37		1	3.0x2.0	15-17	4	2	2
EG05	Portugal	81 clones	RCB	10x43	5	2x43		1	3.0x2.0	18-20	4	2	2
EG06A	Australia	102/0/0/102	RIB	77x84	9	~30x14	117	5x1	2.0x4.0	45	2		1
EG06B	Australia	179/0/0/179	RIB	154x52	9	~35x25	117	5x1	2.0x4.0	10	1		1
EG06C	Australia	114/0/0/114	RIB	78x64	9	~30x16	117	5x1	2.0x4.0	16	2		1
EG07	Australia	594/0/0/594	RIB	204x40	5	~40x40	125	2x1	2.5x4.0	37-64	13	2	3
EG08	Australia	122/0/0/122	RIB	40x174	9	~15x40	117	1x5	4.0x2.0	37	1		1
EG09	Australia	225/0/0/225	RIB[9]	75x112	10	25x32	189	1x4	3.0x2.0	14-16	6	2	
EG10A	Australia	110/0/0/110	RIB	170x75	5	~20x50	55	10x1	2.0x4.0	66-70	2		1
EG10B	Australia	179/0/0/179	RIB	80x191	9	~20x40	117	5x1	3.5x2.5	56	2		1
EG10C	Australia	101/0/0/101	RIB	90x77	9	~15x50	117	5x1	2.0x4.0	49	2		1
EG10D	Australia	114/0/0/114	RIB	65x123	9	~20x40	117	5x1	3.0x3.0	51	2		1
EG11	Australia	101/0/0/101	RIB	80x59	9	~24x25	117	5x1	4.0x3.0	23	2		1
EG12	Australia	110/0/0/110	RRC	106x125	20	16x21		1x3	3.5x2.4	58	1		1
EG13A	Australia	409/0/0/409	RRC	170x76	4	85x24		5x1	2.0x4.0	56	1		1
EG13B	Australia	253/0/0/352	RRC[10]	220x16	2	110x16		5x1	2.0x4.0	24	1		1
EG14A	Australia	408/0/0/408	RRC[10]	130x24	2	95x24		5x1	2.0x4.0	21	1		1
EG14B	Australia	418/0/0/418	RRC[10]	380x22	4	95x22		5x1	2.0x4.0	29	1		1
EG15	Australia	352/0/0/352	RRC[10]	200x32	4	110x16		5x1	2.0x4.0	7	1		1
PA01	Denmark	146/0/1/146	CR	96x72	1-11	-		2x2	2.0x2.0	62-66	2	1	1
PA02	Denmark	37/0/0/37	RCB[9]	100x15	5	20x15		4x1	1.5x1.5	0-3	4	1	1
PA03	Denmark	50/0/0/50	RCB[9]	144x18	4	36x18		4x2	1.6x1.5	7-9	4	1	1
PA04	Denmark	86 clones	RCB	60x24	10	12x12		1	1.5x1.5	28-31	3	1	1
PA05	Denmark	0/51/1/51	RCB[9]	72x20	5	12x20		4x1	1.5x1.5	2-48	4	1	2
PA06	Denmark	0/80/1/66	RCB[9]	80x23	5	16x23		4x1	1.5x1.5	25-25	2	1	1
PA07	Denmark	0/80/1/66	RCB[9]	96x23	6	~13x23		4x1	1.5x1.5	33-68	4	1	1
PA08	Denmark	115 clones	RCB	65x26	10	113x13		1	1.5x1.5	21-26	3	1	1
PA09	Denmark	0/38/1/50	RCB[9]	48x40	5	12x20		4x1	1.5x1.5	50-75	5		2
PM01	Canada	397/0/0/397	RCB	200x60	10	20x60		3nc	0.9x0.6	3-36	8	8	
PP01	Portugal	46/0/1/46	RCB	80x40	8	20x20		4x2	2.0x2.0	0-54	8	3	2

Continued on next page.

Table 7-1 continued.

Code[1]	Country	Pedigree[2]	Des-	Size[4]	Reps		Blks	Plots[5]	Spacing	Fill	Variables[8]		
		OP/CP/Chk/P	ign[3]		n	Size[4]			(m)[6]	(%)[7]	n	Ht	Diam
PP02	Portugal	46/0/1/46	RCB	24x160	8	20x20		4x2	2.0x2.0	20-40	7	3	1
PP03	Portugal	46/0/1/46	RCB	60x60	8	20x20		4x2	2.0x2.0	19-60	7	3	1
PR01A	Australia	27/0/1/27	TS[11]	35x36	8	-		5x1	2.1x2.1	21-22	8	1	2
PR01B	Australia	27/0/1/27	RCB[11]	40x32	8	10x16		5x1	2.1x2.1	0-13	9	1	2
PR02	Australia	309/0/1/309	TS[9,11]	64x50	1-2	-		1x5	2.5x2.5	2-7	8	4	4
PR03	Australia	44/0/1/44	TS[11]	35x80	5	-		5x2	2.5x2.5	1-16	14	7	6
PR04	Australia	430/0/0/430	IB	100x80	4	-	92	1x4	2.1x2.1	10-13	4		2
PR05A	Australia	19/16/2/38	RCB	19x84	6	19x14		1x7	3.0x3.0	20-25	4		1
PR05B	Australia	0/45/2/18	RCB[9]	59x42	6	20x21		1x7	3.0x3.0	24-28	6	1	1
PR06A	Australia	20/0/1/20	RCB[11]	12x20	2	12x10		5nc	4.2x4.2	1-2	3	1	2
PR06B	Australia	20/0/1/20	RCB[11]	20x24	4	10x12		5nc	3.0x3.0	1-2	3	1	2
PR06C	Australia	20/0/1/20	RCB[11]	20x24	4	10x12		5nc	3.0x3.0	4-8	3	1	2
PR06D	Australia	20/0/1/20	RCB[11]	24x40	8	12x10		5nc	2.1x2.1	1-3	3	1	2
PR07A	Australia	26/0/2/26	RCB[11]	20x15	2	10x15		5nc	4.2x4.2	1	1		1
PR07B	Australia	25/0/2/25	RCB[11]	40x30	4	20x15		5nc	3.0x3.0	1	1		1
PR07C	Australia	23/0/2/23	RCB[11]	30x40	8	15x10		5nc	2.1x2.1	1	1		1
PR08A	Australia	4/43/2/47	RCB	72x24	30	12x5		1	3.0x3.0	38	3	1	
PR08B	Australia	48/0/1/71	RCB	60x29	30	12x5		1	3.0x3.0	33	3	1	
PR08C	Australia	47/0/2/37	RCB	60x24	20	10x4		1	3.0x3.0	28	3	1	
PR09A	Australia	19/55/2/76	RCB	60x59	30	10x10		1	3.0x3.0	35	3	1	
PR09B	Australia	110/3/3/116	RCB	70x69	30	14x10		1	3.0x3.0	21-33	3	1	
PS01	Denmark	14/0/2/14	RCB	68x20	5	~13x20		4x4	2.0x2.0	23-57	7	2	1

[1] EG: *Eucalyptus globulus*; PA: *Picea abies*; PM: *Picea mariana*; PP: *Pinus pinaster*; PR: *Pinus radiata*; PS: *Picea sitchensis*. Codes with an alphabetic suffix indicate neighbouring trials on the same site.
[2] No. of families and parents (OP/CP/Chk/P), or clones. OP: open pollinated; CP: control pollinated; Chk: check lots; P: parents.
[3] Designs: CR: Completely Randomised; TS: Trend Surface-completely randomized but with over-replication of check lots; RCB: Randomised Complete Block; RIB: Resolvable Incomplete Block; IB: Incomplete Block; RRC: Resolvable Row-Column.
[4] No. of rows by no. of columns for filled rectangular grid of data.
5 Plot size: no. of rows by no. of columns; 1 for single tree plots; nc for non-contiguous plots with the number of trees per replicate.
[6] Distance between rows by distance between columns.
[7] Range of the proportion of missing values in the rectangular grid of data for the data analysed.
[8] No. of variables; n: total; Ht: Height measurements; Diam: Diameter measurements.
[9] Split plot trial.
[10] Trial blocks non-contiguous.
[11] Check lots over-replicated.

Table 7-2 Data analysed (other than growth) grouped by type.

Trait Type	Variable	Units	Trial(s)	Age	Mean(s)
Bark	Bark thickness as proportion	%	EG01	4	7.8
	of DBH		EG07	4	11.9
Branch	Brach quality score	1-6	PP01-03	12	3.1,2.8,2.7
			PR01A&B	7	3.3,3.2
	Branch node count 1-6m	$\sqrt{n}$	PR01A&B	7	2.7
Deformity	Stem deformation	1-6	PR05A	2	3.0
			PR05B	3,4	3.2,3.4
			PR08A-C	3	3.4,3.7,3.5
			PR09A&B	3	3.5,3.4
Drought	Drought damage	1-9	EG10B-D	5	5.4,9.0,4.4
			EG10A	6	5.6,
Form	Stem straightness	1-4	EG01,02	4,3	2.8,2.5
		1-6	PP01-03	12	3.0,2.8,2.7
		1-6	PR01A&B	7	2.9,3.2
		1-6	PR04	6	3.5
		1-8	PS01	15	5.0
Health	Resistance to needle cast	1-9	PA02	11,14	6.8,7.6
			PA03	12,16	6.6,5.9
			PA04	11	6.4
	Health	1-9	PA08	11	6.6
	Defoliation by fungus	$\sin^{-1}\sqrt{p}$	PR05A	3,4	0.30,0.93
	Dothistroma pinii		PR05B	3,4	0.30,0.90
			PR08A-C	3	0.68.0.64,0.70
			PR09A&B	3	0.81,0.88
	Defoliation by aphid	$\sin^{-1}\sqrt{p}$	PS01	12	0.37
	Elatobium abietinum			18	0.28
	Damage by fungus	1-10	EG03	2	5.92
	Mycosphaerella spp.				

Continued on next page.

Table 7-2 Continued.

Trait Type	Variable	Units	Trial(s)	Age	Mean(s)
Health (cont)	Defoliated by sawfly *Perga affinus*	$\sin^{-1}\sqrt{p}$	EG07	5	0.08
	Egg masses of caterpillar *Mnesampela privata*	$\sqrt{n}$	EG09	3	0.46
	Larval masses of caterpillar *Mnesampela privata*	$\sqrt{n}$	EG09	3	0.27
Leaves	Proportion adult foliage	$\sin^{-1}\sqrt{p}$	EG02	2	0.73
			EG07	2,4	0.60,0.92
	Adult foliage score	1-5	EG09	3	1.3
	Leaf basal lobe length (BASE)	mm	EG07	2	7.7
	Leaf length (LL)	mm	EG07	2	75
	Leaf width (LW)	mm	EG07	2	40
	Length to widest point (LWP)	mm	EG07	2	13
Stems	No. stems at 1.3m	$\sqrt{n}$	EG06A,C	6	1.02,1.02
			EG11	5	1.05
			PR01B	7	1.06
			PR03	6	1.02
			PR04	6,10	1.01,1.01
	No. forks	$\sqrt{n}$	PR01A&B	7	0.25,0.24
			PR04	6	0.30
	No. ramicorns	$\sqrt{n}$	PR01A&B	7	0.75,0.72
			PR04	6	0.78
Wood	Pilodyn penetration	mm	EG01	4	12.6
			PA05	18	22.2
			PA07	10,16	18.9,21.6
			PA09	10,19	19.6,21.8
			PP01-3	12	30.6,32.0,33.0
			PS01	15	14.5
	Spiral grain	°	PA07	16	2.8
			PA09	19	2.7

Trial maps and knowledge of within plot tree numbering systems were used to allocated trees in the trials to row and column positions on a rectangular grid. A complete grid of data was used for computational reasons, so empty tree positions inside the grid, as well as missing or unmeasured trees, were filled with missing values. The complete grids ranged in overall size from 240 to 15280 tree positions (Table 7-1). The missing values comprised between 1 and 75% of the complete grid, so that between 235 and 11645 trees were used for any given analysis. Most trials comprised a single contiguous, or almost contiguous, rectangular grid. Three trials were in physically separated blocks, for which separate grids were established.

Most trials were of open-pollinated or control-pollinated family trials, with between 16 and 594 families from between 14 and 594 parents (Table 7-1). There were four clonal trials with between 57 and 115 clones represented. Plots had between one and sixteen trees representing a particular treatment in each replicate. Eight trials had non-contiguous plots (Libby and Cockerham 1980) of either three or five trees.

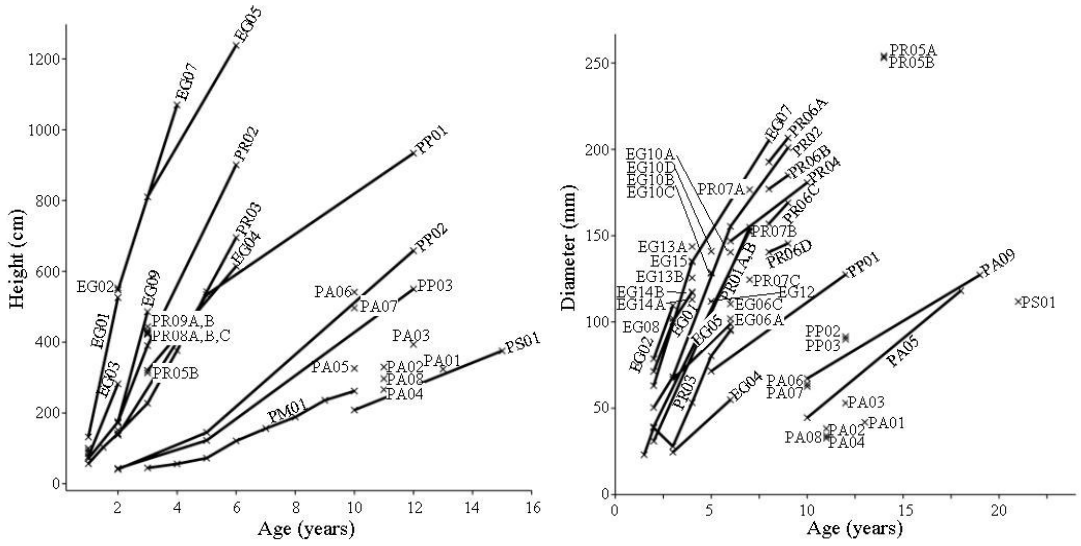


Figure 7-1 Height and diameter (0.3 or 1.3m) growth for each trial.

The trials had up to eight measurements for either height or diameter, with diameter usually measured at breast height (1.3m), but in two instances at 30cm (Table 7-1). The trees at measurement were up to 12m tall, 25cm in diameter, and 21 years of age (Figure 7-1). There were more diameter than height measurements for larger trees because of the difficulty of measuring height on tall trees. Trials of *Pinus radiata* and *Eucalyptus globulus* grew much faster than those of spruce (*Picea spp*.) or *Pinus*

pinaster. The other traits were grouped into 9 types, which covered a wide range of visually scored (mostly form, health and adult foliage), counted (stems, branches and insect abundance) and measured traits (Table 7-2). Health traits encompassed both measures of biotic damage and of the abundance of causal agents. The trait called deformity was a score of stem deformation associated with high nutrient levels on fertilised ex-pasture sites of *Pinus radiata* in north-west Victoria, Australia (Pederick *et al.* 1984). It was treated as a separate trait type, as was drought damage (Dutkowski 1995). The square root of counts was analysed, as was the arcsine-square root of proportion data where a wide range of proportions were present.

7.3.2 Model fitting

The individual tree data for each variable were analysed using a linear mixed model of the general form:

[7-1] $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$

where **y** is the vector of data, **b** is a vector of fixed effects with its design matrix **X**, **u** is a vector of random effects with its design matrix **Z**, and **e** is a vector of residuals. Fixed and random effect solutions were obtained by solving the mixed model equations (Henderson 1984):

[7-2] $$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

where **R** is the variance-covariance matrix of the residuals and **G** is the direct sum of the variance-covariance matrices of each of the random effects which included genetic and experimental design effects. The left hand matrix is known as the coefficient matrix and its inverse is used to calculate the prediction error variances of the solutions.

Two models were fitted to each data set: the design model, where the environmental effects are modelled with only the experimental design features and an independent error, and the spatial model, where an auto-regressive spatial component is added to the design model, as recommended by Costa e Silva *et al.* (2001) and Dutkowski *et al.* (2002).

The spatial component was modelled as separable first-order autoregressive processes in rows and columns with the form

[7-3] σ^2_ξ **[**$\boldsymbol{\Sigma}(\rho_{col}) \otimes \boldsymbol{\Sigma}(\rho_{row})$**]**

where σ^2_ξ is the spatial variance, $\otimes$ is the Kronecker product and $\boldsymbol{\Sigma}(\rho)$ is a first-order auto-regressive correlation matrix with auto-correlation ρ. For a random factor spatially ordered in one dimension with n levels, $\boldsymbol{\Sigma}(\rho)$ has the form:

[7-4] $$\Sigma(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \vdots \\ \rho^2 & \rho & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \cdots & \cdots & \cdots & 1 \end{bmatrix}$$

For the design model, **R** is defined as $\sigma^2_e \mathbf{I_d}$ (where $\mathbf{I_d}$ is an identity matrix of size d, the number of observations). For the spatial model, the spatial component was either included in **R**, or as a random effect in **G**.

For both models, all design features appropriate for the design above the plot level, were fitted as follows:

1) CR & TS: None.

2) RCB: replicates.

3) RIB: replicates and incomplete blocks.

4) IB: incomplete blocks.

5) RRC: replicates, plot-row within replicate and plot-column within replicate.

For trials with contiguous multiple-tree plots, a random plot effect was included, as was the main plot term for split-plot trials. For trials where check lots formed their own main plot, an extra main plot term was fitted for the check lots. All design features were fitted as random effects and permitted to take negative variances (Nelder 1965). The more complex RIB and RRC designs were also analysed with a

simplified RCB design model to look at the relative fit of spatial analysis compared to extra within replicate design features.

Other effects fitted to both the design and spatial models (where appropriate) were:

1) A fixed genetic group effect to account for any differences in the origin of the parents.

2) For family trials, an individual tree random additive genetic effect, with **G** incorporating the numerator relationship matrix (Henderson 1976) to model the genetic covariance between relatives. This allows the simultaneous estimation of breeding values for both the trees in the trials and their parents.

3) For control-pollinated family trials, a random family effect to reflect specific combining ability.

4) For clonal trials, a random clonal (total) genetic effect, without partitioning into additive and non-additive components as pedigree information was not available.

5) For trials with check lots of unknown parentage, an extra independent variance within these check lots. This term was fitted to avoid bias in estimates of additive genetic variances caused by the inclusion of check trees as unrelated base trees in the numerator relationship matrix when their actual relationship was unknown. As check trees made up a substantial proportion of the data in some trials, and could contribute to the estimation of environmental effects, they were retained in the data.

6) For trials with non-contiguous blocks, a fixed block effect was fitted, and the error variances and auto-correlations were constrained to be equal across blocks.

7) Missing values were fitted as fixed effects. This is done for computational reasons as the inverse of an auto-regressive matrix is sparse (tri-diagonal) if it is of full size.

Random effects (other than the main genetic and design terms) with non-significant ($P>0.05$) variances were eliminated from the design model. Their significance was judged using the REML log-likelihood (LogL) and a one or two-tailed Likelihood

Ratio Test (LRT), depending on whether zero was a boundary value (Stram and Lee 1994).

For the spatial model, the starting independent error variance was set at 5 times the spatial variance, the auto-correlations were set at 0.9, and the spatial component incorporated into **R**. Where the spatial model did not readily converge, a number of strategies were used to achieve convergence. The parameter updates between iteration were reduced in size, the design feature variances were constrained to be positive, or eliminated, and lower starting auto-correlations were tried. Alternatively, the spatial component was incorporated into **G**. In common with Saenz-Romero *et al.* (2001) preliminary analyses found a bimodal likelihood profile for some diameter measurements, but with peaks at auto-correlations above 0.9 and around –0.1, indicating the presence of both trend and competition. Where the variogram from the design model indicated the possibility of competition for growth traits (Stringer and Cullis 2002), starting values for auto-correlations of –0.1 were tried in the appropriate direction. In all cases convergence was achieved using one or more of these strategies. Reduced isotropic models were not tested.

Our previous work indicated that explicit modelling of global trend, as recommended by Gilmour *et al.* (1997a), was generally not profitable as very little extra predicted selection gain was achieved. The highest gain was achieved after accounting for a putative extraneous assessment page effect, which was detected by a high auto-correlation in only one direction. When identified by the spatial pattern, the extraneous effects were fitted in extended design and spatial models to test hypotheses about their possible origins, and to see whether extra gain could be achieved.

Best linear unbiased predictions (E-BLUPs) of parent and offspring breeding values, or total genetic values of clones, were obtained from the solutions of the mixed model equations [7-2] using the estimated variance parameters. For each parent, clone or offspring in the trial, the accuracy of the genetic value estimates (the correlation between the true - g - and predicted - $\hat{g}$ - genetic values) was calculated as:

[7-5] $$r_{g\hat{g}} = \sqrt{1 - \frac{PEV}{\hat{\sigma}_g^2}}$$

where PEV is the predicted error variance and $\hat{\sigma}_g^2$ is the estimated genetic (additive or clonal) variance (Quaas *et al.* 1984). The accuracy increases asymptotically toward one as the reliability of the genetic values increases. The variance parameters were estimated by REML and the solutions estimated using the ASReml software (Gilmour *et al.* 1999). Examples of the code to generate the models can be found in the software manual.

Model comparisons

For each variable, the significance of the improvement achieved by using spatial analysis was assessed using a LRT with three degrees of freedom. This is a conservative test as it involves testing one variance, for which zero is a boundary value, and two auto-correlations, for which it is not. Additionally, the spatial variance and auto-correlations must both be non-zero for the spatial component to be meaningful. If the auto-correlations are zero then it becomes another independent error term, and if the spatial variance becomes zero, then again it is another independent error term. Thus the proper test would most likely use less than three degrees of freedom. As the appropriate degrees of freedom is unknown, a conservative test using three degrees of freedom was used.

For variables with significant improvement due to adding the spatial terms, the variances were scaled to the design model error variance and the models were compared by:

- The changes in the spatial, error and genetic variances.
- The sum of the absolute values of the design feature variances.
- The change in accuracy of the predicted genetic values.
- Spearman correlations between the predicted genetic values.
- Relative genetic gains from selection were estimated as the difference in gain of the most desirable 20% of parents (and clones) and 5% of offspring selected by each model on the values estimated from the spatial model.

For the significant spatial models, the spatial auto-correlations were examined for their relationship with tree size. Unusual patterns of spatial parameters were further examined by plotting the spatial residuals and design effects in plan position in order to better understand the patterns of variation.

7.4 Results

There were large differences between traits in their response to spatial analysis (Table 7-3). Stem counts and deformity each had a high proportion of non-significant improvement (>50%) and form and branch characteristics were spread across the improvement classes. Bark thickness, leaf characteristics and drought damage each had significant improvement in every case, however they were drawn from only a few sites. Height performed similarly to diameter, although it gave significant improvement in every case and showed a higher proportion of very large improvements. This comparison is partially confounded by the trials involved, as many height measurements came from the largest trial (PM01) which also had the largest LogL improvements. Comparison of LogL improvements from the 31 instances where a trial had height and diameter measurements at the same age showed similar improvements for both traits, except for about 20% of cases where the LogL improvement for diameter was much smaller (<20%) (not shown).

Large trials where the design terms explained a large amount of variation tended to give a larger improvement in LogL (Table 7-4). A large proportion of non-significant improvements were for small trials (less than 2500 trees) where there was little spatial variation present (the sum of the design feature variances for the design model was less than 10% of the error variance). Additionally, for large trials (>2500 trees) the largest improvements almost all occurred where the sum of the design feature variances was more than 10% of the error variance. This suggests that the spatial model will give larger improvements for larger trials where there is a higher proportion of structured spatial variation which is detected by the design features. However, where design features do not reflect the spatial pattern present, gains may also be made.

For trials with enhanced designs (RIB or RRC) there was usually an improvement in

LogL for all trait types from the simplified design (RCB) model to the design model (using all within-replicate design features) (Figure 7-2). Where the improvement was significant (ΔLogL>1.34 for RIB), there was always a greater improvement from the simplified design model to the spatial model. There was a linear relationship between the improvements which indicates that, where there is within-replicate variation that is being accounted for by the extra design features, this variation is better accounted for by the spatial model. There was no effect of trait type or design, with RRC trials showing the same tendency as RIB trials. In two instances the improvement with the spatial model was greater than expected from the overall relationship. For EG10A Drought-6 there was local trend which was smaller than the 10-tree row-plots, and for EG07 Sawfly-5 there were strips of defoliation along the planting row which were longer than the 2-tree row-plots.

Table 7-3 Distribution of changes in log likelihood with spatial model by trait type.

Probabilities (p) are from a 3-df LRT of the difference between the design and spatial models. ΔLL indicates the change in log-likelihood. n is the number of cases.

		p							ΔLL
Trait Type	n	>0.05	<0.05	$<10^{-2}$	$<10^{-3}$	$<10^{-4}$	$<10^{-5}$	$<10^{-6}$	>50
Bark	2								100%
Branch	8	25%	13%	25%		25%	13%		
Deformity	8	63%	13%			13%		13%	
Diameter	72	8%	4%	8%	8%	8%	1%	42%	19%
Drought	4							25%	75%
Form	9	22%	22%	22%		11%		11%	11%
Health	22	5%	9%			9%	9%	36%	32%
Height	63		8%	2%	3%		10%	25%	52%
Leaves	8			13%	25%			63%	
Stems	9	78%	22%						
Wood	11	36%						64%	

Table 7-4 Distribution of changes in log-likelihood with respect to the ratio of the sum of design feature variances to the error variance in the design model and the number of observations.

Probabilities (p) are from a 3-df LRT of the difference between the design and spatial models. ΔLL indicates the change in log-likelihood.

no. values	<2500						>2500					
Design Variance	<10%	10-20%	20-30%	30-40%	40-50%	>50%	<10%	10-20%	20-30%	30-40%	40-50%	>50%
p>0.05	7.4%	2.8%	0.9%				1.4%					
p<0.05	4.6%	0.9%	0.5%		0.5%		0.9%					
$p<10^{-2}$	2.8%	0.9%	0.5%					1.4%				
$p<10^{-3}$	2.8%				0.5%			1.4%				
$p<10^{-4}$	1.9%	2.3%	0.5%	0.5%				0.5%				
$p<10^{-5}$	0.5%	1.9%	0.5%	0.5%				1.4%				
$p<10^{-6}$	3.2%	4.2%	3.7%	1.9%	0.9%	3.2%	3.7%	6.5%	3.2%	0.9%		0.5%
ΔLL>50	0.9%	0.5%		0.5%	0.9%	2.3%	1.4%	6.5%	3.2%	2.8%	2.3%	6.5%

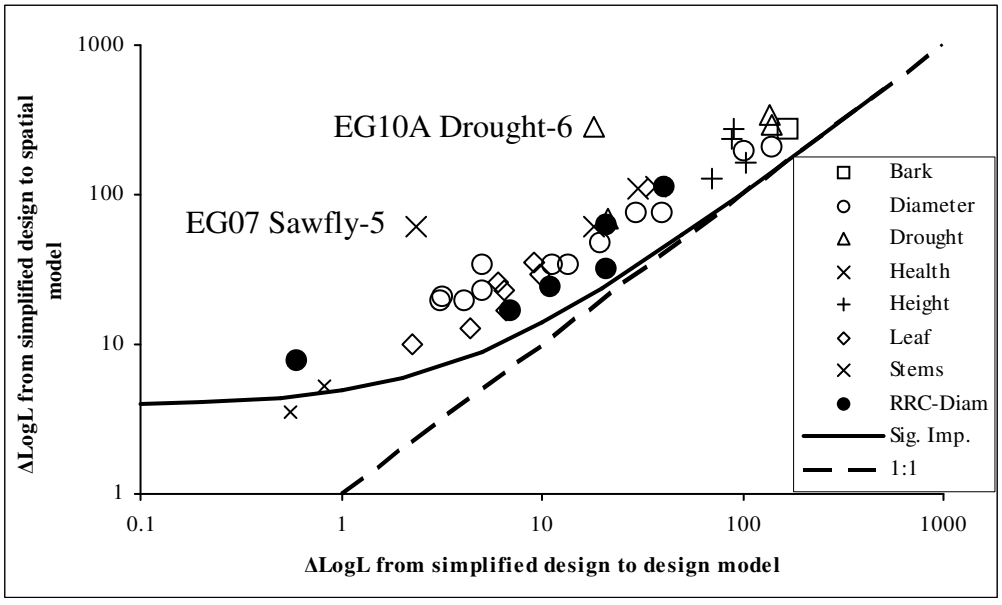


Figure 7-2 Changes in log likelihood (ΔLogL) from adding within replicate design features and spatial components to the simplified design model for more complex trial designs.

Most trials are RIB designs - RRC designs are indicated. Sig. Imp. denotes significant improvement (p<0.05) of the spatial model over the design model according to a 2-tailed LRT.

The sum of the spatial and independent error variances with the spatial model was usually greater than the design model error variance (Figure 7-3). The independent error variance was always reduced with the spatial model, but in most cases did not decline by more than 40%. The spatial variance was usually less than the independent error variance, indicating the dominance of random error. There were two distinct groups which did not follow these trends. The first exhibited large decreases in the independent error variance and was associated with relatively low auto-correlations indicating patchiness, or no independent error variance at all for small negative auto-correlations, indicating competition, and early height at PM01. The second showed small (<20%) decreases in the error variance but very high spatial variance. It was associated with diameter measurements with very high auto-correlations (0.99 and above on at least one axis) representing strong global trends (large patch sizes). It is a characteristic of the auto-regressive process that the nominal variance for a given phenotypic variance increases with the auto-correlation.

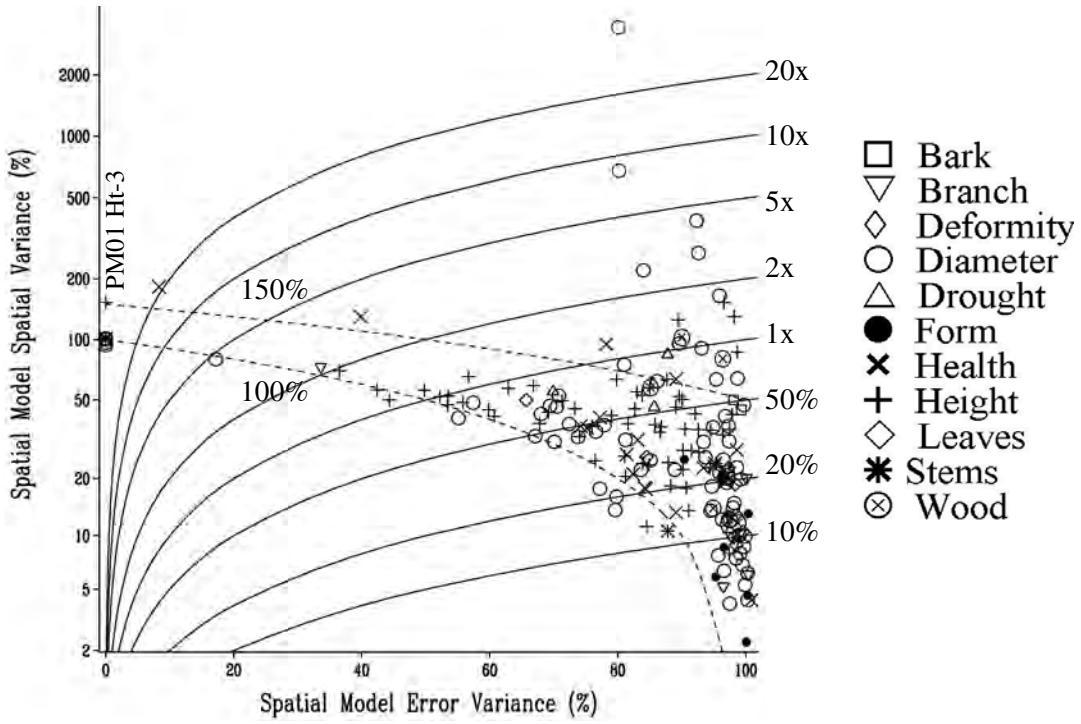


Figure 7-3 Spatial and independent error variance in spatial model by trait type.

The variances are expressed as a proportion of the design model error variance. The solid lines show the ratio of spatial to error variance, and the dotted lines show their sum. Results from spatial models with non-significant model improvement are not shown.

The auto-correlations were predominantly symmetric and high (> 0.8) (Figure 7-4). As the spatial variance was small, the net auto-correlation of residuals was much lower. The auto-correlation indicates the patchiness of the spatial surface. The surfaces were thus mainly characterised by smooth global trend (large patches). Lower auto-correlations (0.5-0.8), indicating more local trend (patchiness), were dominated by health traits and growth measurements from PR03. Both stem count variables with significant improvement had low or negative auto-correlations.

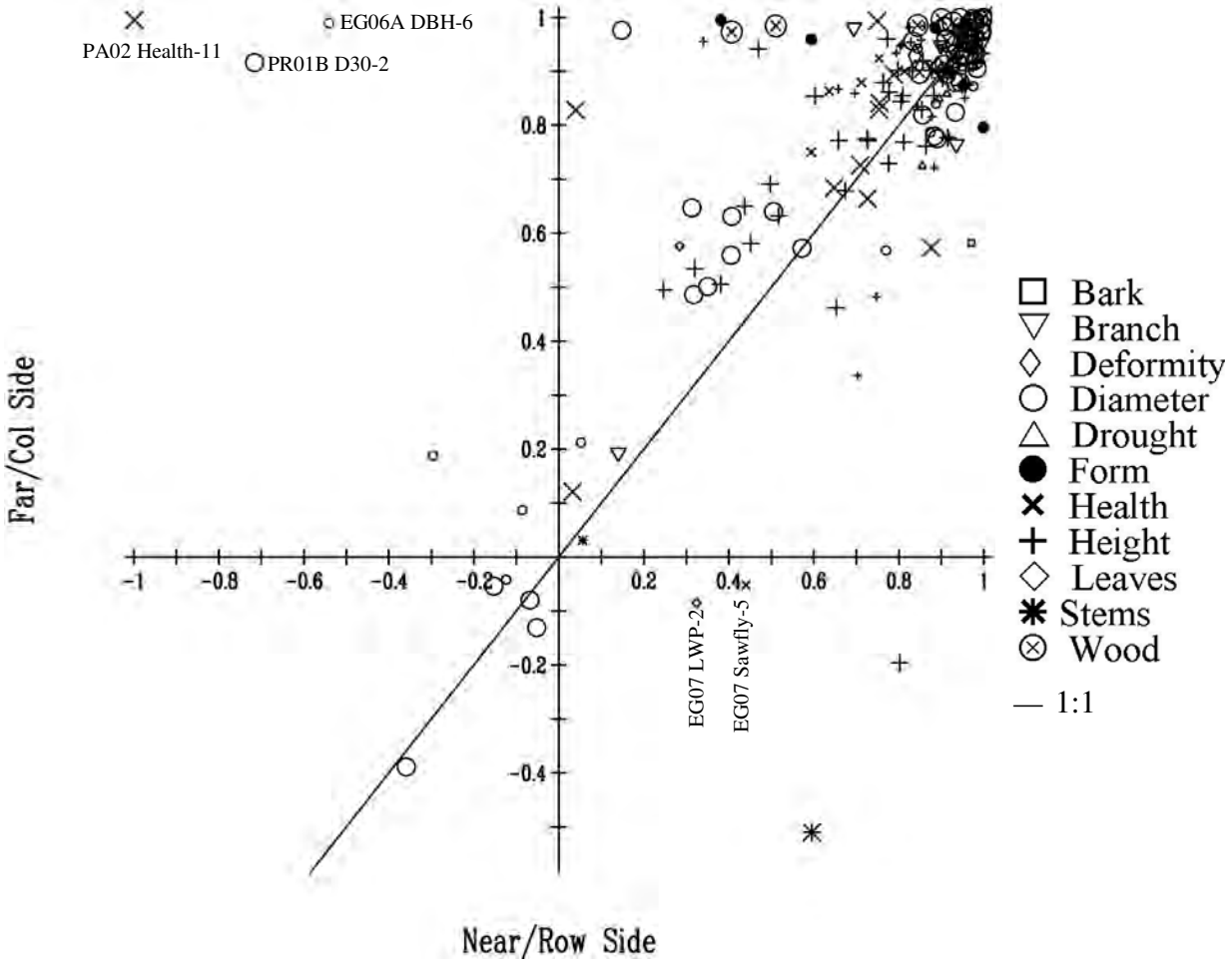
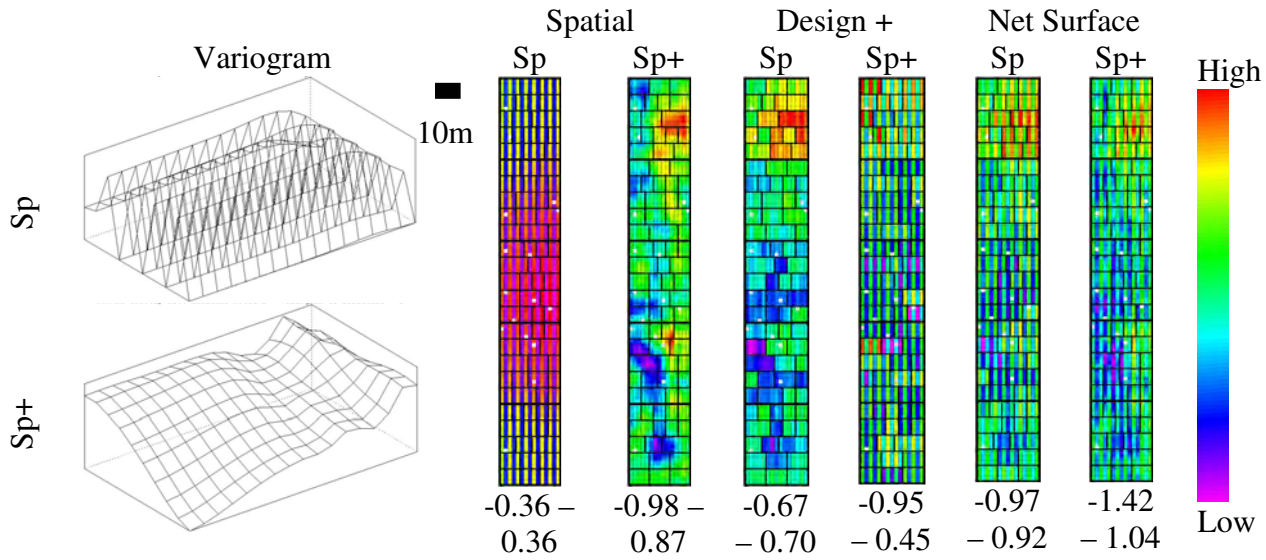
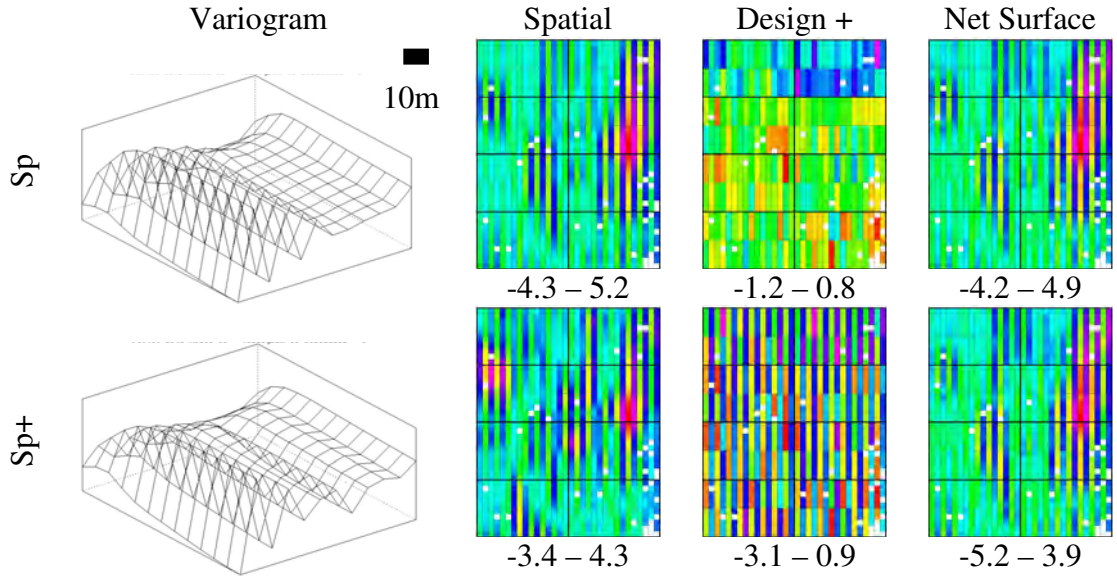


Figure 7-4 Distribution of auto-correlations by trial shape and trait type.

Large symbols represent trials with square spacing for which the row and column auto-correlations are shown. Small symbols represent rectangular spacing for which the near and far tree spacing auto-correlations are shown. Results from spatial models with non-significant model improvement are not shown. Data points mentioned in the text are labelled with the trial code, trait name and age.



(a) PA02 Health-11 (1-9)



(b) PR01B D30cm-2 (mm)

Figure 7-5 Spatial (Sp) and extended spatial (Sp+) models for variables with highly asymmetric auto-correlations.

The variogram and maps of the spatial surface (Spatial), the design and extraneous effects (Design+) fitted, and the sum of the two (Net Surface). The lines represent the design feature boundaries and the numbers represent the range of values.

Continued on next page.

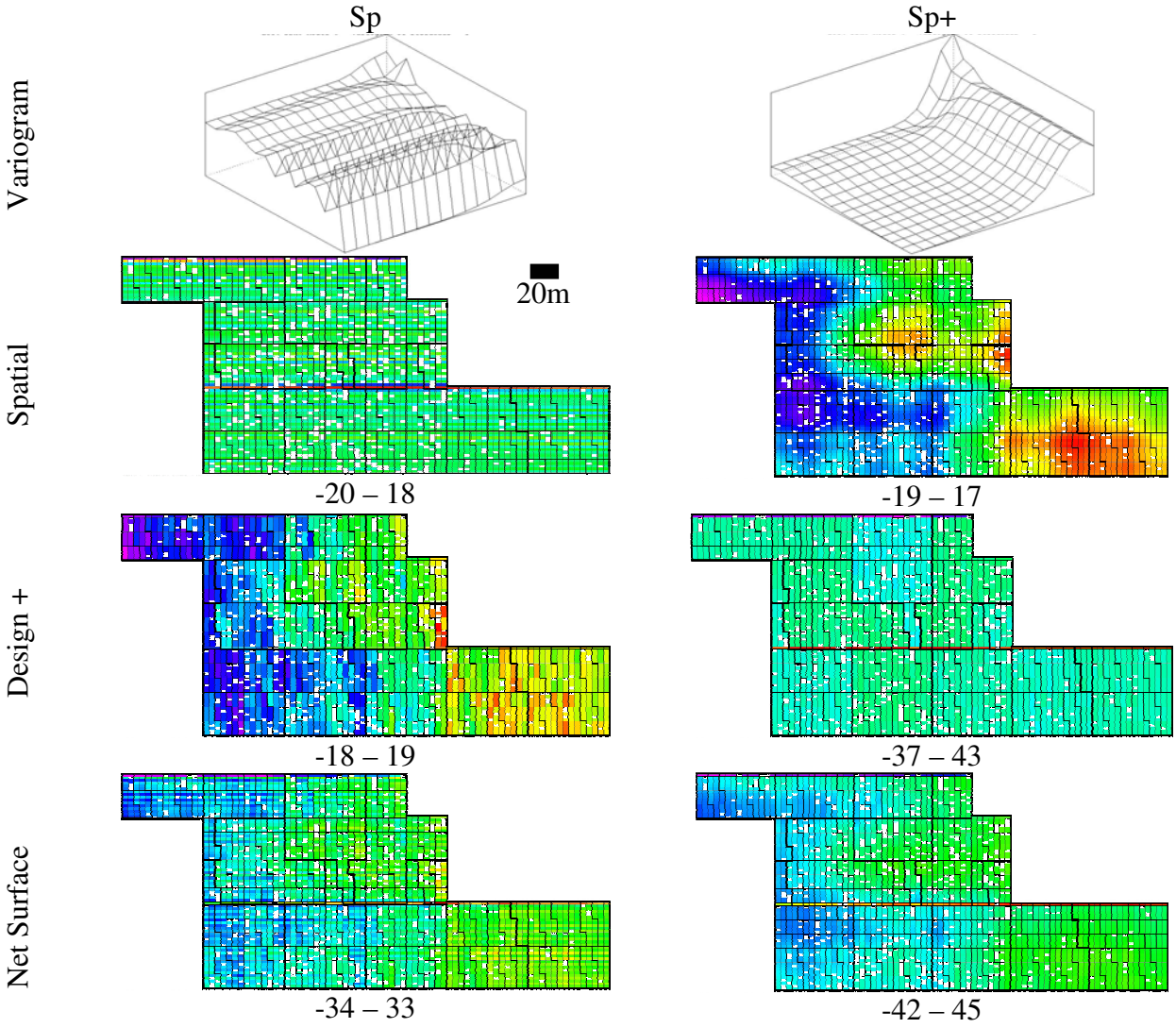


Figure 7-5 **continued.** (c) EG06A DBH-6 (mm)

A group of four square-planted trials with large diameters (>17cm) gave negative auto-correlations on both axes, indicating strong competition. Three smaller (DBH 9-13cm) trials with rectangular planting also gave negative auto-correlations on the near side, with positive, or less negative, auto-correlations on the far side. There were no cases of negative-auto-correlation for height, however, the trees measured for height were generally smaller than those measured for diameter because of the difficulty of measuring height on large trees. For cases where height and diameter were measured at the same age, there were no systematic differences in the auto-correlations.

While the majority of the auto-correlations were more or less symmetric, there were a

number of cases of marked asymmetry. Higher auto-correlations in the direction of assessment across a number of trait types where the trials were assessed across the whole length or width of the trial suggests assessment direction effects, or serial auto-correlation in assessor effects. This suggestion is supported by the high proportion of variables based on subjective scores (form and health) or the repeatability in use of a measurement device (pilodyn penetration or bark thickness). Health at age 11 in PA02 showed a strong negative auto-correlation between adjacent assessment columns, suggesting an effect of assessment direction (Figure 7-5a). Fitting this as a fixed effect showed it to be highly significant with the spatial variance increasing 40-fold and the auto-correlations becoming symmetric and moderate. Diameter at 30cm above ground level in PR01B, showed similar asymmetry suggesting a directional assessment effect, perhaps related to consistency of measurement height in each direction (Figure 7-5b). Measurement direction was significant as a fixed effect when added to the design model, but became much less significant and did not substantially change the spatial components of the spatial model. Neither of these effects was apparent in later measurements of similar traits on each trial, supporting the suggestion of transient assessor effects. Diameter at EG06A is a special case: the negative auto-correlation in one direction seems to be related to 2 lines of control seedlots planted at right angles to the normal plots with markedly faster growth in one instance and poorer in another (Figure 7-5c). These large differences from adjacent rows are being manifested as a negative auto-correlation. Fitting each line as a fixed effect was highly significant for both the design and spatial models. With the spatial model, it increased the spatial variance and made the auto-correlations symmetric. Adding random row or columns effects showed substantial improvements over the design model in all five other instances tested of marked asymmetry with one auto-correlation close to one. In no case, however, was the model as good as the spatial model. In only the case of groups of columns consistent with pages of the assessment sheets for form at EG02 did adding these extra factors improve the spatial model, as previously reported in Dutkowski *et al.* (2002).

For the instances where the spatial model improved over the design model, the spatial model usually dramatically decreased the sum of the variance components of the design features (Figure 7-6), with the spatial variance increasing accordingly (not

shown). In three-quarters of these instances, the reduction was over 50%. There were some cases of total elimination of the design feature variances, however in all but one instance this was where the design variances had been constrained to be positive to achieve convergence. The proportion of cases where one or more design feature variances was negative increased from 10% for the design model to 36% for the spatial model. The design feature variances sum increased with the spatial model in only 9% of cases. These cases were usually associated either with competition, or single-tree plot RCB trials where the replicates spanned the trial and the dominant trend was at right angles to the blocking. The plot variance for trials with multiple-tree plots within larger design features declined from a median of 9% of the design model error variance to 4% with the spatial model, in line with the overall reduction of design feature variances. However, in line with our previous work, the proportion of the sum of the design feature variances due to plots increased from a median of 50% to 69%.

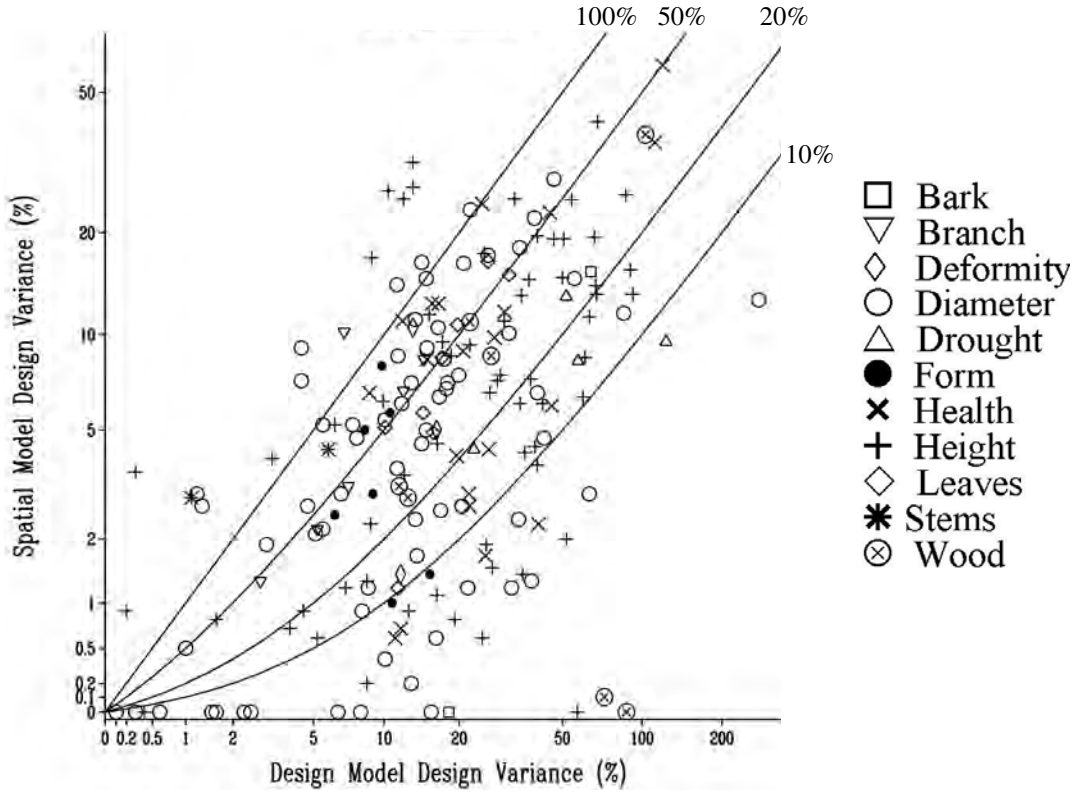


Figure 7-6 Effect of spatial analysis on variance due to design factors by trait type.

The sum of the design factor variances is expressed as a proportion of the design model error variance. The lines show the proportion of reduction with the spatial model. Results from

spatial models with non-significant model improvement are not shown.

In most cases, the genetic (additive or clonal) variance did not change much between the design and spatial models (Figure 7-7). The genetic variances both increased and decreased in value, with the largest proportional changes occurring at low values. The unit heritability for PM01 for early height did not change, nor did the zero value for seven variables. In two instances the additive variance increased from 0, to 2.5% for PP03 and 5% for PS01, but in 2 other instances it also increased for PS01, as previously reported for that trial in Costa e Silva *et al.* (2001). In a number of instances the additive variance was markedly reduced (eg PP01-DBH12) and for PR01C early height it was eliminated. The accuracy of additive genetic values for both parents and offspring generally increased only a small amount (<0.05), in line with the overall small reduction of the error variance (Figure 7-8). The larger increases and decreases were associated with similar changes in the additive variance.

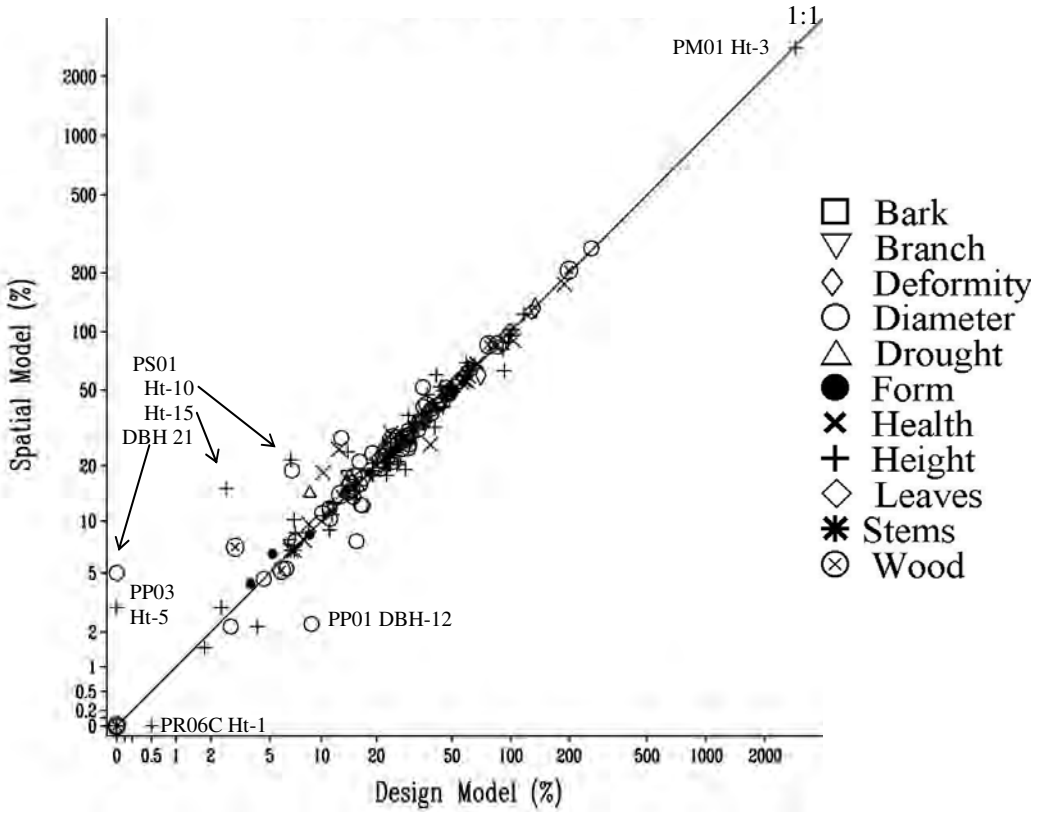


Figure 7-7 Genetic variance (additive or clonal) for design and spatial models by trait type.

The variances are expressed as a proportion of the design model error variance. Results from spatial models with non-significant model improvement are not shown. Data points

mentioned in the text are labelled with the trial code, trait name and age.

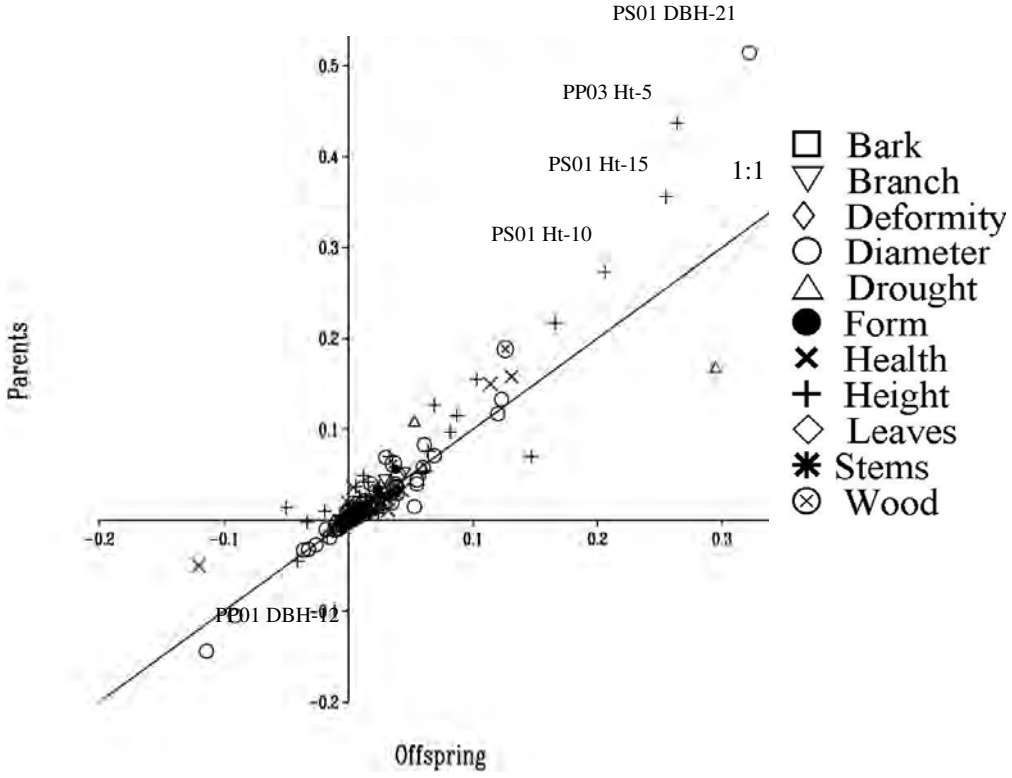


Figure 7-8 Change in additive genetic value accuracy for spatial models by trait type.

Results from spatial models with non-significant model improvement are not shown. Data points mentioned in the text are labelled with the trial code, trait name and age.

The correlation between the predicted breeding values from the two models was usually greater than 0.9 for both parents and offspring (Figure 7-9). The correlation was usually similar for both parents and offspring. The lowest correlations (<0.8 for parents) were associated with one of the TS trials (PR02), 2 of the 3 large plot *Pinus pinaster* trials (PP01& PP03) and one of the form measurements (PS01). These same trials, together with PM01 heights and EG10A drought, dominated the next group of low correlations.

Selection on the better spatial model increased relative genetic gain in the main by less than 5%, with the median gain being 1.3% for parents and 2.1% for offspring (Figure 7-10). Despite the different selection intensities and variation in the number of offspring per parent, parental and offspring gains were broadly correlated. However, for a group of trials with a small (<50) number of parents, there was often no change in the parents selected (and thus no parental gain) while there was still

offspring gain. This accounts for the lower median gain for parents, but this is offset at high gains by parental gains tending to be larger. As expected, the same group of trials which gave low correlations also tended to dominate the instances of high gain.

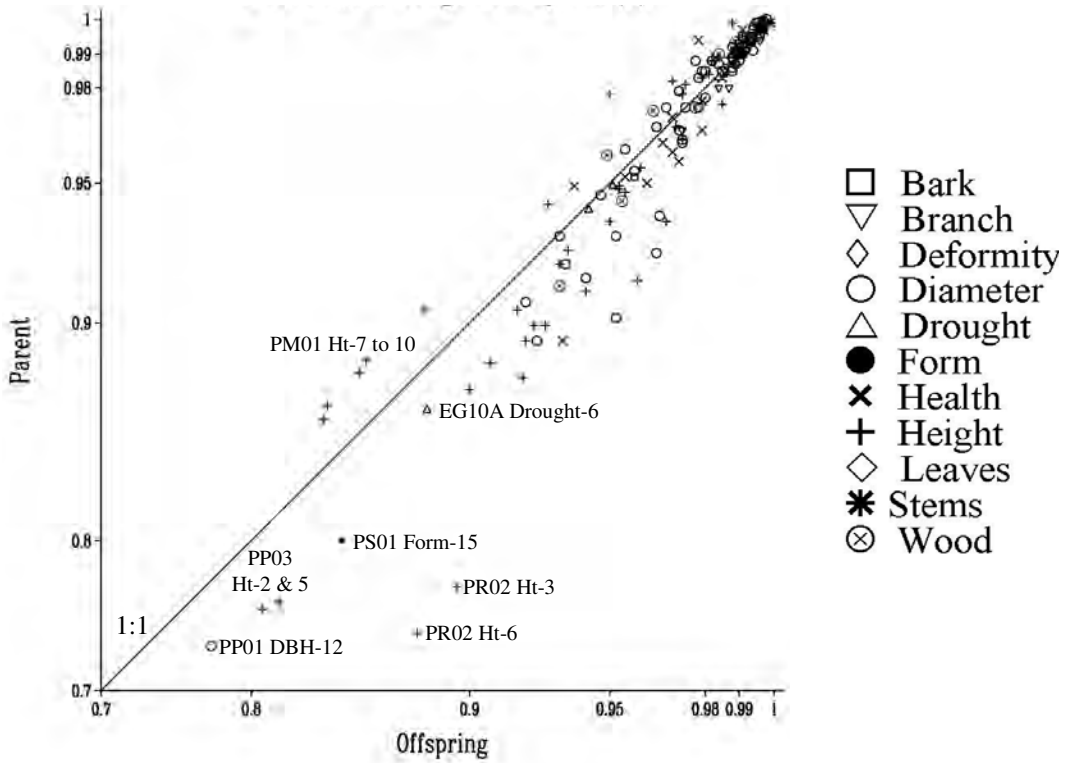


Figure 7-9 Correlation between breeding values from the design and spatial models for parents and offspring by trait type.

Results from spatial models with non-significant model improvement are not shown. Data points mentioned in the text are labelled with the trial code, trait name and age.

Height accounted for most of the 10% of cases where both parent and offspring gains exceeded 10%. The three health traits with high gains were unusual as they had high correlations of breeding values between the two models, but disproportionately large gains due to skewed breeding values. Diameter gave high gain only in two instances. Where both height and diameter were measured at the same age, there was a tendency for higher gains for height, especially for parents (not shown). All the high gains for height came from only four trials, indicating a common effect across tree sizes for those trials with the design features not accounting for the spatial patterns. Similarly, of the 15 trials with gain greater than 10% for parents or offspring, seven had more than one trait with such high gains. Five of these came from RCB designs, but it included two of the four completely randomised designs. All three of the large eight

tree plot trials (PP series) were in this group, as was PM01 which had 1200 trees per replicate and no within replicate blocking features. The PP series and PR01B also had less than 50 parents, which may have contributed to their relatively high parental selection improvements.

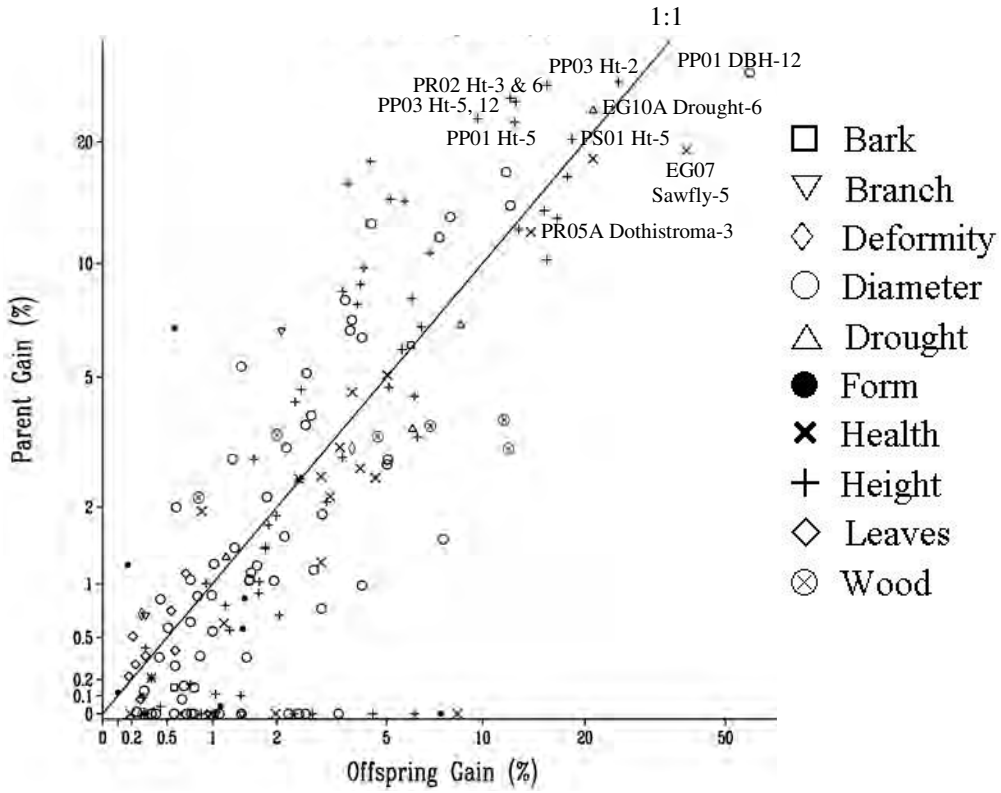


Figure 7-10 Gain from selection for spatial models by trait type.

Gain is for selection of the best 5% of offspring and 20% of parents. Gain is expressed as the relative value on the scale of genetic values from the spatial model of the trees selected with the spatial model compared to those selected by the design model. Results from spatial models with non-significant model improvement are not shown. Data points mentioned in the text are labelled with the trial code, trait name and age.

Gains with the extended spatial model were achieved in some instances where the model could be successfully fitted. Health in PA02 gave the highest genetic gain (4.3% for offspring gain and 1.3% for parents), with column groups for form at EG02 giving 1.6% for offspring and 0.2 for parents, and the lowest gain for EG06A DBH (0.7% for offspring and 0% for parents).

Most of the results which stand out from the general trends came from a few trials whose results stood out in several ways. The height measurements at PM01 (the largest trial) showed the largest LogL increases, high selection gains and, from

age 6 years onwards, most of the cases where the design features variances sum increased with the spatial model. It was also strange in that the first measurement had a heritability close to one, which was presumably due to some sort of residual family propagation effect, as the heritability was reduced thereafter. The auto-correlations were markedly asymmetric at this age, but at all ages showed a stronger, but less marked, auto-correlation between rows. This trial was unusual in its design as it used non-contiguous plots. Due to its long narrow shape, the replicates spanned the width of the trial, however major trends appeared as stripes along the long axis of the trial, which crossed the replicate boundaries.

The *Pinus pinaster* trials (PP01-3), with their large plots of eight trees, were characterized by high selection gains, with the growth measurements from all three trials being in the group of five trials which dominated parental selection gains of greater than 10%. Diameter at age 12 in PP01 showed marked edge effects (as reported for age 5 in Dutkowski *et al.* 2002), which resulted in very high auto-correlations in both directions and the highest spatial variance. This data also showed the highest overall gain for both parents (28%) and offspring (59%), but the additive variance decreased by 80% with a concomitant decrease in breeding value accuracy. With the design model, the high breeding value trees were predominantly along the edges of the trial, whereas with the spatial model, they were more uniformly distributed through the trial.

Two of the trend surface designs (PR01A and PR02) also had relatively high selection gains. For PR01A, the gains were higher than for the adjacent RCB design trial (PR01B) with the same genetic material.

7.5 Discussion

Spatial analysis should be routinely applied to forest genetic trials. These results, and our previous work using the same methods (Costa e Silva *et al.* 2001 and Dutkowski *et al.* 2002), show that in the majority of cases the augmentation of the design model with a spatial component gives a statistically improved model. This is reasonable since design features do not usually represent identifiable sources of variation except as an approximation to spatial variation which is expected to occur. While the

relative genetic gains from selection are generally modest, in certain instances they can be large, without the cost of extra replication or testing. A lack of model improvement can generally be ascribed to a lack of spatially structured environmental variation, which to a degree reflects the nature of the traits measured. The spatial model generally provides a more realistic and satisfying description of the pattern of variation on a site, and can lead to a much better understanding of the nature of that variation.

The majority of genetic trial measurements are of growth and, in more than nine out of ten cases across a range of tree sizes, these will respond to spatial analysis. Other traits can respond as well, and the response is related to the amount and type of spatially structured variation present for the trait at the specific trial and how well the experimental design matches it. While this makes generalisations difficult, across all our work we have found that spatially structured variation in the presence of inadequate blocking, large plots, and poor orientation of blocks with regards to dominant trends, all increase the chance of large model improvement, and thus selection gain, following spatial analysis. We have found that, with experience, the best indicator of the likely response is a visual inspection of the data plotted in its spatial position, bearing in mind the plot size used. If patches or trend are seen, then the data will usually respond to spatial analysis, unless the patches or trends were specifically anticipated in the design. The production of such a map is predicated on the allocation of spatial positions to all trees, which is often a barrier to spatial analysis in large or historic forestry trials. If such a map can be produced, it is only a small amount of extra work to actually test the spatial model.

The auto-regressive model has been found to be robust in a number of situations (Grondona *et al.* 1996). We also use it because it is computationally efficient (its inverse is sparse), but other models are possible. The linear variance model of Williams *et al.* (2005) and the auto-regressive models are similar in terms of computational efficiency and the auto-correlation pattern is similar at the high auto-correlations we have found. Applied within replicates, the two models have been shown to be similarly statistically efficient and valid (Azais *et al.* 1998). Dutkowski *et al.* (2002), however, found that restriction of the auto-correlation to within replicates gave a poorer fit where competition was not present. Applying our

method to the Williams *et al.* data with a RRC plus long-column design model, showed a marginal model improvement (ΔLL=6.4) but a remarkably small decrease (1.1) if the auto-correlation is restricted to within replicates. Their data showed a very strong long column effect due a very poor edge row, and the long columns effectively gave the across replicate linkage that the continuous auto-regressive error model would otherwise give.

Costa e Silva *et al.* (2001) and Dutkowski *et al.* (2002) indicated that separation of global and local trends was unnecessary, and in this work we have generally not tried to do so. Brownie and Gompertz (1997) reported undesirable consequences where global trend is masking local trend. The auto-regressive model is very flexible in the surface it fits, and while it is classically used as a local trend model, it will often fit smooth global trend better than alternative global trend models such as trend surfaces, or splines. Convergence problems can arise if both terms are included as they are often trying to model the same effects. It is sometimes possible to fit both simultaneously, but with little change in the net surface. The spatial variation is simply shuffled between the two effects, with the auto-correlation (and thus notional patch size) changing as alternative global models are fitted. Fu *et al.* (1999) and Joyce *et al.* (2002) used row and column effects to model global trend and estimated local patch sizes in the residuals using a variogram. This patch size is entirely dependent on the efficacy of the global trend model. Adding random row and column terms to the design model for the data of Joyce *et al.* (2002) (PM01) confirmed that these global trend terms were significant. Adding the auto-regressive terms, however, greatly reduced these terms as the effects were better fitted by the auto-regressive terms. Saenz-Romero *et al.* (2001) used a trend surface to model global trend and an isotropic exponential error auto-correlation model for local trend. In one dimension, the latter is equivalent to the auto-regressive model (Cullis *et al.* 1998). Reanalysing their data, we found that we were indeed able to fit both local and global terms. However, in accordance with our previous experience, the net surfaces were very similar.

Where a combination of local and global trend exists, then retaining the design features in the spatial model ensures, to a degree, that trend at both scales is fitted. The design features will account for trend at a size commensurate with their size.

Whatever sort of trend the auto-regressive model effectively fits, if trend exists at other scales, this can be picked up by the appropriately sized design features, if they exist. The high auto-correlation parameters found indicate large patches traversing multiple plots (global trend) are being modelled. The rise in the proportion of design variance due to plots with the spatial model indicates that the plots are picking up small patches (local trend). Maps of the net surface supported this observation, but allowing the design feature variances to be negative was necessary for this to happen. The maintenance of plot variance when there is high auto-correlation may indicate when significant local trend exists, and that it may be worth fitting.

While we have found fitting of smooth global trend usually fruitless, there is some evidence that fitting extraneous effects, such as the assessor effect we found, could be profitable. Instances where we did find extra gain in modelling the extraneous variation were however quite rare and extreme in their auto-correlation asymmetry. More work is needed to confirm if modelling extraneous variation is most profitable only in this circumstance. It would seem reasonable that the auto-regressive component would be best at modelling the widespread continuous trends that we found, rather than discontinuous effects like assessor effects. We suggest that the diagnostic tools recommended by Gilmour *et al.* (1997a) and colour intensity maps of the spatial surface be used to detect such extraneous effects. All terms should however be fitted in a single model, rather than in a multi-stage process, so as to judge as well as possible the need for them all, although this may not be always possible if different terms are trying to fit the same effect. The use of a single stage approach, where the spatial, extraneous and treatment effects are estimated simultaneously, should be superior, even with similar models, as it will avoid problems of confusing spatial variation with local random aggregations of good or poor families.

Not surprisingly, our data confirmed the theoretical benefits of incomplete block designs already demonstrated by Fu (2003) in forest trees and many workers in agricultural variety trials. More blocking factors of different sizes should generally allow better accounting for spatial variation at a variety of scales. This is in accord with the general recommendation that patchy variation (low to moderate auto-correlations) needs smaller design features to cope with it. Unfortunately the spatial auto-correlation is rarely known before genetic trials are planted, and the sampling

of soil parameters to assess likely auto-correlations (Fagroud and Van Meirvenne 2002) would not generally be feasible. The generally symmetric auto-correlations that we have found supports the recommendation that blocks be as square as possible (Correll and Cellier 1987). Conversely, if there are strong row or column effects expected (giving a higher auto-correlation in one direction) then appropriate long blocks are necessary to best approximate the high auto-correlation at large distances. In contrast to Correll and Cellier (1987), we found such effects to be relatively rare.

Despite the benefits of incomplete block designs, in common with Kempton *et al.* (1994), Baird and Mead (1991), and others, we found that the spatial model was an improvement over incomplete block analysis. We also confirmed the observation of Baird and Mead (1991) that, where incomplete block models were better than randomised complete block models, the spatial analysis was even better. The smooth nature of the auto-regressive surface should generally be better that a discontinuous block based one, but the more blocking factors used, the better. We would always recommend that the best possible design is used as good design will avoid confounding of treatment and error effects. It provides a strong basis for any analysis, and as we recommend the retention of design features in the model to complement the auto-regressive terms, you should use the best design with multiple blocking factors of various sizes. Where strong row or column effects are expected, appropriate row-column designs would be critical. The new class of spatially balanced designs (Azais *et al.* 1998; Williams *et al.* 2005) should be considered. Azais *et al.* (1998) showed that neighbour designs can improve efficiency, but that there can be problems with some systematic designs.

We generally found symmetric auto-correlations, which would allow a reduced isotropic model to be used. The instances of anisotropy we found, however, justify considering an anisotropic model, even if isotropy is expected. Many of the cases of asymmetric auto-correlation found in our data suggest the presence of assessment effects. These have only been detected because of the alignment of assessment order with rows or columns. Their detection in both measured and subjectively scored traits suggests that special care needs to be taken in the way in which assessments are carried out to minimize such effects. Additionally, models including assessor identity

and allowing for serial auto-correlation within assessors could be used (Diggle *et al.* 1994).

The small number of negative auto-correlations found indicated that competition was not dominant in our data set. To a degree, this was to be expected as much of our data was for small trees which may not have entered a competitive stage of stand development, and some of the older trials had been thinned. Nevertheless, our data was broadly representative of the measurements carried out in progeny trials. Fox *et al.* (2001), in their review of a number of forestry studies, confirmed the pattern of auto-correlation proposed by Reed and Burkhart (1985) with positive values before canopy closure, sometimes negative values afterwards, and a re-emergence of positive values after thinning. They did, however, note that for many studies trend was generally considered to be dominant and competitive effects were often not found. Our data similarly suggests that trend dominates. The trials we analysed are also large compared to the plots of up to 0.1 ha used by Reed and Burkhart (1985), or only the nearest neighbours examined by Huhn and Langner (1995). This meant that trends were likely to be more important and boost auto-correlations in comparison with those studies. While we found that trend was dominant, the bimodal nature of the likelihood surface indicates that it may well be present in more cases, just not dominant. Our spatial model can account for trend using the design features and competition using the auto-regressive structure, however it is unlikely to be the best model. Competition will probably act at the phenotypic rather than error level (Besag and Kempton 1986; Stringer and Cullis 2002). Durban *et al.* (2001) and Stringer *et al.* (2002) found competition common in plot data from sugar beet and sugar cane trials, respectively, and, like (Magnussen 1994) for forest trees, they have proposed models for the separation of trend from competitive effects. Resende *et al.* (2004) used the Stringer *et al.* (2002) approach on two forest tree data sets and were able to model both spatial trend and competition. However, in both data sets there were negative auto-correlations for a simple auto-regressive model, which indicates that there was strong competition present, a situation that we have found to be relatively rare at the ages of assessment we used.

A number of the trials examined used non-contiguous plots which, as they did not represent any coherent spatial unit, we were reluctant to use. In common with

Loo-Dinkins and Tauer (1987), we usually found close to zero variance for such plots in any case. Non-contiguous plots have the advantage of avoiding missing plots in plot level analysis, allowing systematic thinning if used with inter-locking designs (Libby and Cockerham 1980), and reducing the environmental contribution to family mean estimates compared to multiple tree plots (Loo-Dinkins and Tauer 1987). If an individual tree mixed model is used, then the missing value problem is eliminated. An equivalent single-tree plot design that would allow systematic thinning should be possible, and would use the maximal efficiency of single-tree plots that Loo-Dinkins and Tauer (1987) found.

7.6 Conclusion

Spatial analysis should be routinely applied in forest genetic trial analysis where the spatial arrangement of trees can be determined. The improvements in model fit are often high, and where they are not, a standard design model can be applied. The improvements in selection due to spatial analysis are generally modest, but can be large. The models and tools are available. While our work has focused on the development of the methods and tools, and demonstrating where gains might be made, we look forward to seeing spatial analysis used as a routine tool in forest genetic trial analysis.

7.7 Acknowledgements

We thank the Arboretum (KVL, Hørsholm, Denmark), Estação Florestal Nacional (Lisbon, Portugal), WA Plantation Resources (Manjimup, Western Australia), the Cooperative Research Centre for Sustainable Production Forestry and Gunns Ltd (Tasmania, Australia), RAIZ (Lisbon, Portugal), ForestrySA (Mount Gambier, South Australia), the Southern Tree Breeding Association (Mount Gambier, South Australia), Hancock Victorian Plantations (Melbourne, Victoria, Australia), the Genetic Resource Management Program (Ontario, Canada) and the Instituto Nacional de Tecnología Agropecuaria (INTA) (Buenos Aires, Argentina), in association with Sociedade Portuguesa de Papel (SOPORCEL) (Portugal), for making the data available. We also thank the Fundação para a Ciência e Tecnologia (Lisbon, Portugal) for their financial support of the second author. The comments of Brian

Cullis greatly helped in improving the models and the paper.

Chapter 8 Concluding Discussion

The race classification for *Eucalyptus globulus* developed here (Chapter 2) has proved to be stable when tested with independent data sets. Dutkowski (2000a) and Lopez *et al*. (2001) applied the same methods to independent native forest seed collections of 438 and 214 *E. globulus* families from 30 and 10 localities with more than 3 families, respectively (Table 8-1). The new collections covered some of the same collection localities and a number of new ones, extending the geographic range and giving more intensive sampling of some areas. While generally a narrower range of production traits were used, principally survival, growth, pilodyn penetration and transition to adult foliage, new traits (form and presence of forks/multiple stems) were also included.

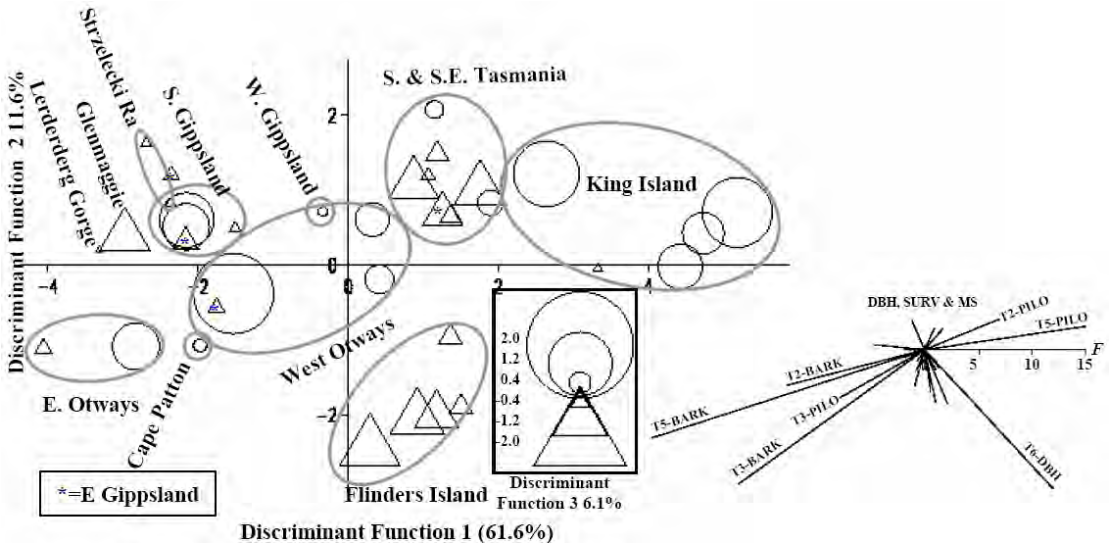


Figure 8-1 *Eucalyptus globulus* locality variation on the major discriminant functions from Dutkowski (2000a).

The vectors on the right show the direction of the trait discriminant coefficients, with the length proportional to the between-locality F-ratio. The vectors are labelled with the trial number (T1-T6), and a trait code which follows Table 2-2 with the addition of MS for presence of multiple stems. Identifiable sub-race groups are shown.

The results from Dutkowski (2000a) show that the previous groupings were largely intact, but showed that the new areas (Lerderderg Gorge and Glenmaggie) were distinct, and that there was a cline from the Strzelecki Ranges to coastal South Gippsland and through the Otway Ranges (Figure 8-1). Poorly sampled East and

West Gippsland localities did not form a coherent group. As in Chapter 2, the major discriminating variables were bark thickness and pilodyn penetration. The clustering from Lopez *et al.* (2001) also confirmed the patterns of variation, but clustered the new Cygnet locality closer to the Southern race than the South-eastern race (Figure 8-2).

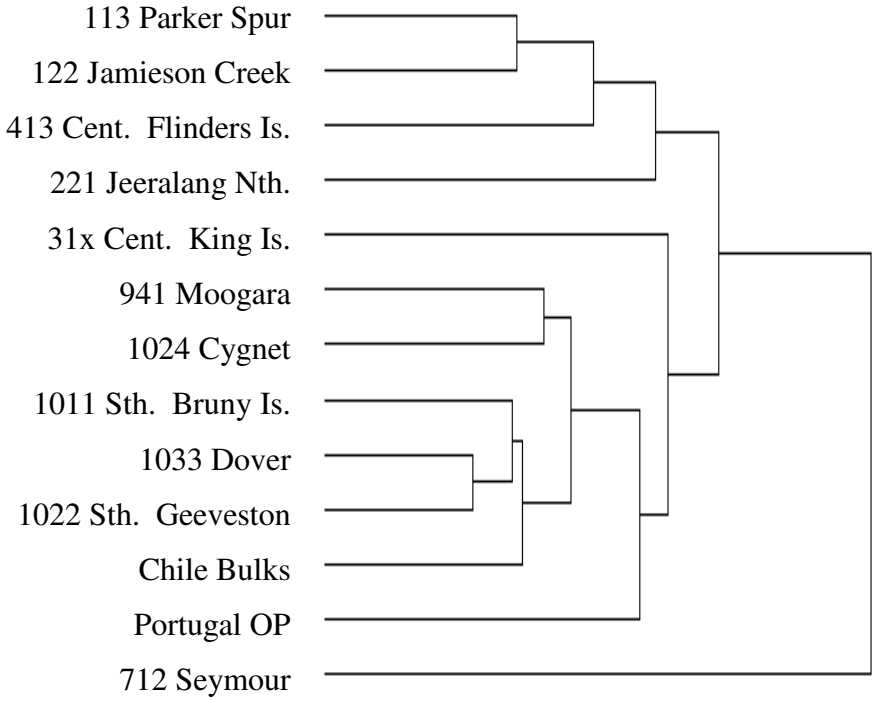


Figure 8-2 UPGMA clustering of *E. globulus* native stand and land race localities from Lopez *et al.* (2001).

The new data sets have allowed the extension of the classification to new areas, and better delineation of boundaries in areas that were previously poorly sampled. New collection localities at Lerderderg Gorge, Glenmaggie and East Gippsland are proposed as separate races as they occur in areas with closer affinities to the other subspecies (Jordan *et al.* 1993) and are separated from each other geographically and somewhat on the discriminant functions (Figure 8-1). New sub-race boundaries have been defined in previously sampled areas (Figure 8-3). The western end of the distribution in the Otway Ranges has shown itself to be somewhat different, and because the area is physiographically distinct, it has been split off as a separate sub-race (Figure 8-3a). The cline between the Strzelecki Ranges and South Gippsland has allowed the delineation of a new sub-race (S. Gippsland Foothills) (Figure 8-3b). The new Cygnet locality has moved that area from the Southern race to the South-eastern

races (Figure 8-3c). The entire classification is available on the World Wide Web at http://members.forestry.crc.org.au/globulus/index.html.

Table 8-1 Localities sampled in revisions of the *E. globulus* race classification.

Code	Name	Latitude	Longitude	1999	2000	2001
11	Lerderderg Gorge	37°32'	144°21'		14	
21	Glenmaggie	37°49'	146°40'		13	
31	Monkey Top	37°26'	148°32'		5	
32	Erinundra	37°28'	148°50'		2	
33	Wiebens Hill	37°37'	148°44'		6	
34	Lind NP	37°37'	149°4'		5	
111	Otways State Forest	38°45'	143°26'	44		
112	Cannan Spur	38°45'	143°30'	21		
113	Parker Spur	38°47'	143°34'	59	38	30
114	S.W. Lavers Hill	38°44'	143°18'	6	7	
115	Skeenes Creek	38°40'	143°43'		10	
121	Lorne P.O.	38°31'	143°58'	18	19	
122	Jamieson Creek	38°36'	143°52'	13	24	21
123	Cape Patton	38°39'	143°48'	21	21	
211	Madalya Road	38°31'	146°30'	9	10	
212	Hedley	38°38'	146°29'	9		
213	Bowden Road	38°25'	146°40'	5	12	
214	Port Franklin	38°40'	146°16'	5	0	
215	Yarram	38°32'	146°39'		16	
216	Mardan	38°27'	146°6'		15	
217	Alberton West	38°36'	146°31'	1	27	
218	Toora North	38°36'	146°21'	3		
219	Won Wron	38°30'	146°43'		13	
221	Jeeralang North	38°20'	146°30'	51	16	30
222	Jeeralang	38°25'	146°31'	3	1	
231	Wilsons Promontory Lighthouse	39°7'	146°25'	16		
311	Central King Island East	39°55'	144°4'	10	16	
312	South King Island East	40°2'	144°2'	1	2	
313	King Island South West	40°3'	143°55'	3		
314	King Island Central West	39°53'	143°55'	9	9	
315	King Island North	39°38'	144°3'		5	
316	Central King Island North	39°46'	143°57'		1	
411	North Flinders Island	39°49'	147°53'	10	15	
412	Central North Flinders Island	39°55'	147°58'	15	9	

Continued on next page.

Table 8-1 Continued

Code	Name	Latitude	Longitude	1999	2000	2001
413	Central Flinders Island	40°3'	147°59'	15	20	38
414	Central East Flinders Island	39°58'	148°10'	1	2	
415	South Flinders Island	40°13'	148°8'	12	8	
421	North Cape Barren Island	40°21'	148°15'	10		
422	West Cape Barren Island	40°23'	148°3'	34		
431	Clarke Island	40°32'	148°10'	6		
511	St. Helens	41°13'	148°15'	13		
521	German Town	41°36'	148°13'	5		10
531	Royal George	41°52'	147°59'	10		
541	Pepper Hill	41°38'	147°49'	10		
611	Badgers Creek	42°0'	145°16'	14		
612	Macquarie Harbour	42°20'	145°19'	8		
613	Little Henty River	41°55'	145°12'	11		
711	Mayfield	42°14'	148°0'	6		
712	Seymour	41°40'	148°17'		10	
721	Taranna	43°4'	147°52'	5		
731	Triabunna	42°28'	147°55'	10		
741	North Maria Island	42°37'	148°5'	7		
811	Port Davey	43°17'	145°54'	6		
911	Ellendale	42°38'	146°43'	5		
921	Mount Dromedary	42°42'	147°7'	4		
931	Collinsvale	42°50'	147°12'	5		
941	Moogara	42°46'	146°53'	26	24	23
942	Leesons Hill	42°50'	146°57'		9	
951	Hobart South	42°55'	147°16'	10		
961	Jericho	42°25'	147°15'	5		
1011	South Bruny Island	43°22'	147°17'	7		10
1020	Geeveston: Between Castle Forbes Bay, Geeveston & Police Point	43°11'	146°57'		15	
1021	North Geeveston	43°8'	146°56'	3	1	
1022	South Geeveston	43°11'	146°53'	7		4
1023	Blue Gum Hill	43°2'	146°53'	4		
1024	Cygnet	43°7'	147°5'		10	23
1031	Recherche Bay	43°30'	146°53'	5		
1032	Strathblane	43°23'	146°58'	1	4	
1033	Dover	43°16'	147°0'	6	13	25

1999 refers to Dutkowski *et al.* (1999), 2000 to Dutkowski (2000a), and 2001 to Lopez *et al.* (2001).

Figure 8-3 Revised race and sub-race boundaries.

Thick lines with bold type indicate races, and thin lines with smaller type indicate sub-races. Dots indicate the centre of collection localities.

The traits used for the original and updated classifications are very broad, and provides a very good basis for such work. Given the divisions already identified, more traits will only lead to further subdivision, or make some clustering clearer. Jordan *et al* (1998) noted that there was little differences between localities once subraces were taken into account for a wide variety of traits in the trials used in the original classification. The only exception was for flowering time, with large locality differences in some subraces. Unfortunately, insufficient flowering time information was available to include in the classification. Economic traits were included in the analysis (growth and pilodyn penetration) and later age measurements of growth, or direct measurements of basic density are likely to be so highly correlated with those measurements already undertaken, that they will provide little discriminatory power. The trails cover a wide range of environements (Tasmania, Western Australia and Argentina), yet the patterns revealed were very similar as growth was not a very discriminating trait. Kraft pulp yield is the other major trait of importance for these species, however its expense means that there is insufficient data to add to the classification. Available data indicates little differences between the races established (Raymond *et al.* 2001), so it unlikely to have much discriminatory power. Further refinement of the classification is likely to come from molecular genetic studies, as all known large collections of *E. globulus* in field trials have already been used. Molecular studies do not rely on common garden experiments to differentiate genetic from environmental effects.

Since the work of Nesbitt *et al.* (1995), there has been much more extensive work on the geographic structure of molecular marker variation (Freeman *et al.* 2001; Jones *et al.* 2001; Steane *et al.* 2001; Jones *et al.* 2002). The latest and most extensive, reported in Potts *et al.* (2004b), has shown a remarkable concordance with this revision of the classification at the race level (Figure 8-4). While there are differences, the molecular groupings make sense geographically, with spatially adjacent races more likely to cluster together on the molecular information. Some races closely related on the molecular phylogeny however show large difference in quantitative traits, so their division is justified. There are still insufficient samples with molecular information at small enough spatial scale to better delineate race or sub-race boundaries.

The patterns of geographic variation at the locality level was examined using synthetic traits derived from the first axis of a principal components analysis of similar measurements on different sites and at different ages. This used locality means derived from family means fitted as fixed effects. Other methods of genotype by environment analysis such as AMMI (Additive Main and Multiplicative Interaction) could have been used to derive overall measures of performance. AMMI suffers from a number of limitations (de Resende and Thompson 2004), not least of which is sensitivity to imbalance in family representation between sites. Factor analytic models (Jennrich and Schluchter 1986; Piepho 1998) can be used in the mixed model context to generate BLUPs of overall family and locality performance using the correlations between sites to obtain loadings for each site on the primary factor. This should weight the data for differences between sites in the number of trees per family, family composition of the localities, and missing localities on some sites. Where there are multiple measurements on a site, such as for growth, an inter-site factor analytic model could be used in conjunction with within site inter-trait or inter-age correlations to avoid over-weighting the results from that site.

The classification has been used extensively for the prediction of breeding values in Australia (D. Pilbeam *pers. comm.*), Chile (R. Sanhueza *pers. comm.*) and Portugal (N. Borralho *pers. comm.*). It has formed the basis of studies of genetic variation in a number of traits such as predicted pulp yield (Raymond *et al.* 2001), attack by the Southern Eucalypt Leaf Beetle (*Chrysophtharta agricola*) (Rapley *et al.* 2004a),

Autumn Gum Moth (*Mnesampela privata*) oviposition preference (Rapley *et al.* 2004b), resistance to marsupial browsers (O'Reilly-Wapstra *et al.* 2001), vulnerability to a seed destroying wasp (*Megastigmus sp.*) (Lorkin *et al.* 2004), flowering control (McGowen *et al.* 2004), *Phoracantha* resistance and drought survival (Toval 2004), resistance to *Gonipterus scutellatus* (Basurco and Toval 2004), *Mycosphaerella* resistance (Potts *et al.* 2004a), rooting ability (Cañas *et al.* 2004), coppicing ability (Whittock *et al.* 2003) and sawn timber traits (Greaves *et al.* 2004).

The race classification of *E. nitens* has also been similarly used for the prediction of breeding values in a number of programs (Dutkowski 2000b; Kube and Dutkowski 2002). There has been no other quantitative analysis to confirm the patterns of variation found in this study. Moran (*pers. comm.*), however, has found in a microsatellite survey of 300 of the families used for the quantitative work that there was no significant geographic structure at the locality level. This contrasts with the population structure found for the species as a whole (Byrne *et al.* 1998; Cook and Ladiges 1998). It is possible that over the much more restricted geographic area of central Victorian part of the distribution of this species that gene flow has been sufficient to make the molecular pattern differ from the selection effects of local adaptation that may only affect a few genes. Adaptive multiple locus trait divergence can, however, occur with limited divergence of allelic frequencies due to linkage disequilibrium (Latta 2004).

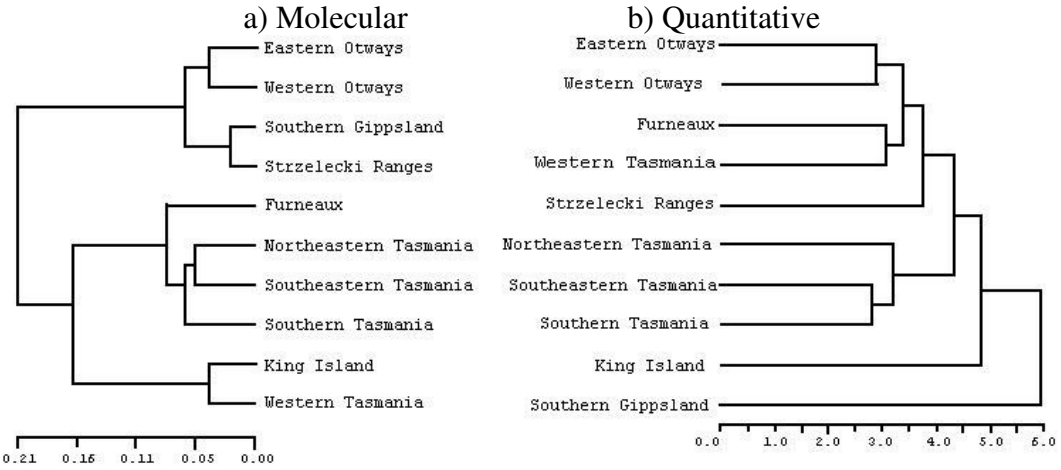


Figure 8-4 Comparison of molecular and quantitative affinities between races.

The UPGMA dendrograms are based on a) Nei's genetic distances derived from an analysis of 8 microsatellite loci using nearly 400 native trees and b) Mahalanobis' distances based on 38 quantitative traits measured on the trial used in Chapter 2. From Potts *et al.* (2004b).

The program for the inversion of the A matrix for parental inbreeding and partial selfing has been used for estimation of variance components in *E. globulus* (Lopez *et al.* 2002), and for a variety of eucalypt species in Western Australia (R. Mazanec *pers. comm.*) using population estimates of partial selfing. It is being incorporated into the next version of the ASReml software (A. Gilmour *pers. comm.*).

However, accounting for partial selfing alone does not account for all the peculiarities of mating systems in native forests. Most notably, mating may occur with related individuals due to family structure in the forest springing from limited seed dispersal and limited pollen dispersal mechanisms (Hardner *et al.* 1998; Skabo *et al.* 1998). While accounting for partial selfing may act as a surrogate for all forms of related mating in the forest, there will be differences in the average relationship between trees due to the precise form of mating. Given the relative insensitivity of breeding values to even large errors in pedigree if the estimate of additive variance is correct, it is unlikely that minor changes due to the mating pattern will be important.

Additionally, the method cannot overcome all of the problems inherent in analysis of partially selfed breeding populations. The method does not account for the effects of differential partial selfing and differential inbreeding depression on both population means and variances. Hardner and Potts (1995) found that selfing reduced growth and increased both within and between family variance when compared with outcrossed material. Borralho and Potts (1996) found reductions in diameter growth in stands of low stocking which Hardner *et al*. (1996b) confirmed was related to differences in outcrossing rate. Burgess *et al.* (1996) found that growth was also related to outcrossing rate in *E. grandis*. Hardner *et al.* (1996a) found that including inbreeding depression was critical in improving prediction of breeding values in a mixed outcrossed and selfed population, although they included dominance in the model and did not modify the additive relationship matrix. Volker (2002) found that open pollinated seedlots yielded higher heritability than, and poor correlation with, control pollinated seedlot performance for growth, but not for pilodyn. Correcting for relatedness is clearly only part of the solution for the problems in breeding value estimation in open pollinated seedlots.

For traits without the confounding effects of inbreeding depression then the patterns

of within and between family variation may enable estimation of mating system parameters. In a mixed population of control and open pollinated families then profile likelihoods could be used to estimates the selfing rate because of the expected effects on the distribution of breeding values and thus phenotypes. Large trials would however be necessary to achieve this due to the variation in variance component estimates that can occur with limited numbers of parents, families and offspring per parent. Work on some small trials of *E. globulus* with mixed mating types (Volker 2002) indicated that there were large differences between traits and trials in estimated selfing rates and these had large standard errors (Dutkowski, unpublished data). Similarly, profile likelihood could be used in mixed base and advanced generation trials, where outcrossing had removed the effect of inbreeding in base population parents, to estimate the degree of that inbreeding. However in addition to the problems generated by small trials, it is unusual to have such pedigrees in the same trial, making such comparisons difficult. Amalgamating data from such pedigrees from different trials would introduce confounding due to site differences on variances.

While there has been an increase in the interest in spatial analysis of forest genetic trials (Kusnandar and Galwey 2000; Mora-Garces and Ramirez 2000; Hamann *et al.* 2001; Saenz-Romero *et al.* 2001; Joyce *et al.* 2002; de Resende *et al.* 2004; Williams *et al.* 2005) it is far from being a routine tool for analysis. Partly this is probably due to the lack of uniformity in approach in the literature, and the general strategy in much of the work of trying to model both global and local trend in a stepwise manner. This makes the analyses more complex and time consuming. Lack of readily available software may also contribute. As the tools are now available, and the approach suggested largely avoids complex stepwise trend modelling, we should now see this approach being more universally adopted.

While the spatial work has been carried out on a trial-by-trial basis, it is possible to simultaneously apply the spatial model to more than one trial. Some of the trials were made up of non-contiguous blocks that could well represent separate trials. Although the parameter estimates were constrained to be the same for all blocks, this is not necessary, nor even desirable, for different sites. Despite the efficiencies in the software used, this approach may still lead to computational problems for very large breeding programs with multiple traits. A similar approach to that used by

Hamann *et al.* (2001) could be used where the estimated spatial surface is simply subtracted from the data for each variable. Applying this method to a large breeding value estimation project with over 60 trials (Dutkowski, unpublished work) has revealed that for each variable the parameter estimates and breeding values are essentially unchanged when compared to the spatial model. Simpler models can then be applied in large across-site analysis with fewer computational constraints.

Other aspects of the adaptations that White and Hodge (1989) indicate are necessary for these models to be applied to large tree breeding evaluation programs have been studied. Costa e Silva *et al.* (2005) found that even in dealing with a sample of even aged clonal trial sites that the variances were heterogenous, making the application of BLUP more difficult across trials. However, they found that ignoring this heterogeneity made little difference to the selections made. Wei and Borralho (1998) showed that thinning could be accounted for in the estimation of variance components when using a multivariate model. Using simulated data, Ye *et al.* (2004) found that accurate breeding value prediction in second generation progeny trials could be undertaken without first generation data. Ye had, however, assumed that first generation trials were balanced and complete, whereas in many instances they are not. Part of the benefit in amalgamating data across generations is to link previously unlinked first generation trials from which selections and crosses have been made to allow more selections from better provenances or trials. As computational limitations are removed, we should be concentrating on more appropriate models, rather than simpler models to keep within computational limitations.

The approach of Ye *et al.* (2004), however, may help in the intractable problems with open-pollinated base generations with differential selfing and expression of inbreeding depression. The base generation growth information could be excluded from analyses to avoid problems of biased variances and poor correlations with control crossed performance. Alternatively base generation growth information could be treated as a separate trait with a reduced heritability and a low correlation with true performance.

Other tree breeding simulations are also starting to use the individual additive model. They have been used for analysis of crossing strategies (Borralho and Dutkowski

1998), sampling strategies (Apiolaza *et al.* 1999; Dutkowski and Raymond 2001; Pilbeam and Dutkowski 2004), and clonal forestry (Dutkowski 2004), as well as in the PopSim (Mullin and Park 1995) software (T. Mullins *pers. comm.*).

Reporting of the use of BLUP with the Numerator Relationship Matrix for routine prediction of individual tree breeding values in large programs has not been widespread in the published literature. These models are being used more and more for large scale evaluations, however these tend to be for proprietary use, rather than for publication. ASReml (Gilmour *et al.* 1999) is being used in a number of programs: New Zealand *Pinus radiata* (L. Apiolaza *pers. comm.*), U.S.A. *Pseudotsuga menziesii* (C. Dean *pers. comm.*), and Chilean (R. Sanhueza and J. Brawner *pers. comm.*) and Portuguese (N. Borralho *pers. comm.*) *E. globulus*. It has been mainly used on a univariate basis, as has SAS for *Pinus taeda* (B. Li *pers. comm.*). Multivariate analysis has been carried out using PEST (Groeneveld 1990) in Australia for *E. globulus* (Jarvis *et al.* 1995), *Araucaria cunninghamii* and a number of *Pinus spp*. and hybrids (M. Dieters *pers. comm.*), TREEPLAN (Kerr *et al.* 2001) for *Pinus radiata* and *E. globulus* (McRae *et al.* 2004), and BioCat (De Veer *et al.* 2001) for *Pinus radiata* in Chile. Use of all data for a whole breeding program needs robust systems for data management as well as analysis. This is another impediment in many breeding programs.

In conclusion, as breeding programs progress and more information across generations accumulate, and more questions about its application to trees are answered, then the benefits and usage of an individual additive model are likely to grow. This thesis has attempted to look at some aspects of such a model and provide some help in its use in specific circumstances. It has derived race classifications for two species and demonstrated how their use can improve models and increase gain. Some of the problems in the use of partially inbred parents and partially selfed seed have been overcome. A simple approach to spatial analysis has been developed and has been shown to give substantial gains in some cases. I hope to see more widespread use of these models in the future.

References

Akaike H. (1973) Information theory and an extension of the maximum principle. In proceedings of '2nd International Symposium on Information Theory'. (Eds B.N. Petrov and F. Czaki) Tsahkadsov, Armenia. pp. 267-281. (Akademia Kiado: Budapest, Hungary)

Almeida M.H. (1993) Geographic variability in *Eucalyptus globulus* Labill. PhD thesis, Universidade Tecnica de Lisboa.

Almeida M.H., Pereira H., Miranda I., Tome M. (1995) Provenance trials of *Eucalyptus globulus* Labill. in Portugal. In proceedings of CRCTHF-IUFRO Conference 'Eucalypt Plantations: Improving Fibre Yield and Quality'. (Eds B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits and C.A. Raymond). 19-24 February, Hobart, Tasmania. pp. 195-198. (CRC for Temperate Hardwood Forestry: Hobart, Tasmania)

Anekonda T.S., Libby W.J. (1996) Effectiveness of nearest-neighbor data adjustment in a clonal test of Redwood. *Silvae Genetica* 45, 46-51.

Apiolaza L.A., Burdon R.D., Garrick D.J. (1999) Effect of univariate sub-sampling on the efficiency of bivariate parameter estimation and selection using half-sib progeny tests. *Forest Genetics* 6, 79-87.

Apiolaza L.A., Garrick D.J. (2001) Analysis of longitudinal data from progeny tests: some multivariate approaches. *Forest Science* 47, 129-140.

Araújo J.A., Lemos L., Ramos A., Ferreira J.G., Borralho N.M.G. (1997) The RAIZ *Eucalyptus globulus* breeding program: A BLUP rolling-front strategy with a mixed clonal and seedling deployment scheme. In proceedings of 'IUFRO Conference on Silviculture and Improvement of Eucalypts'. (Eds A.R. Higa, E. Schaitza and S. Gaiad). 24-29 August, Salvador, Brazil. Vol. 1: Tree improvement strategies. pp. 371-376. (EMBRAPA: Colombo, Brazil)

Askew G.R., El-Kassaby Y.A. (1994) Estimation of relationship coefficients among progeny derived from wind-pollinated orchard seeds. *Theoretical and Applied Genetics* 88, 267-272.

Azais J.M., Monod H., Bailey R.A. (1998) The influence of design on validity and efficiency of neighbour methods. *Biometrics* 54, 1374-1387.

Baird D., Mead R. (1991) The empirical effciency and validity of two neighbour models. *Biometrics* 47, 1473-1487.

Bartlett M.S. (1978) Nearest neighbour models in the analysis of field experiments. *Journal of the Royal Statistical Society, Series B* 40, 147-174.

Basford K.E., Williams E.R., Cullis B.R., Gilmour A.R., Hammer G.L. (1996) Experimental design and analysis for variety trials. In 'Plant adaptation and crop

improvement.' (Ed. M. Cooper) pp. 125-138. (CAB International: Wallingford, UK)

Basurco F., Toval G. (2004) Estimated genetic parameters and breeding values for resistance against *Gonipterus scutellatus* in 13-year-old *Eucalyptus globulus* in Northwest Spain. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 155-158. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Besag J., Kempton R. (1986) Statistical analysis of field experiments using neighbouring plots. *Biometrics* 42, 231-251.

Borralho N.M.G. (1994) Heterogeneous selfing rates and dominance effects in estimating heritabilities from open pollinated progeny. *Canadian Journal of Forest Research* 24, 1079-1082.

Borralho N.M.G. (1995) The impact of individual tree mixed models (BLUP) in tree breeding strategies. In proceedings of CRCTHF-IUFRO Conference 'Eucalypt Plantations: Improving Fibre Yield and Quality'. (Eds B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits and C.A. Raymond). 19-24 February, Hobart, Tasmania. pp. 141-145. (CRC for Temperate Hardwood Forestry: Hobart, Tasmania)

Borralho N.M.G., Cotterill P. (1994) Genetic improvement of *Eucalyptuis globulus* for pulp production. In proceedings of '*Eucalyptus* for Biomass Production'. (Eds J.S. Pereira and H. Pereira) Lisbon, Portugal. pp. 86-89. (Commission of the European Communities)

Borralho N.M.G., Dutkowski G.W. (1998) Comparison of rolling front and discrete generation breeding strategies for trees. *Canadian Journal of Forest Research* 28, 987-993.

Borralho N.M.G., Potts B.M. (1996) Accounting for native stand characteristics in genetic evaluations of open pollinated progeny from *Eucalyptus globulus* base population. *New Forests* 11, 53-64.

Box G.E.P., Jenkins G.M. (1970) 'Time Series Analysis, Forecasting and Control.' (Holden-Day: San Francisco)

Braysher B.D., Dear B.S., Boschma S., Cullis B.R., Sandral G.A. (2001) Improving the efficiency of pasture breeding programs through the use of spatial analysis. In proceedings of 'Tenth Australian Agronomy Conference'. 29 January-1 February, Hobart, Australia. (Australian Society of Agronomy)

Brown M.J., Bayly-Stark H.J. (1979) 'The vegetation of Maria Island.' Wildlife Division Technical Report 79/1. (National Parks and Wildlife Service: Hobart, Tasmania)

Brownie C., Gumpertz M.L. (1997) Validity of spatial analyses for large field trials. *Journal of Agricultural, Biological and Environmental Statistics* 2, 1-23.

Burgess I.P., Williams E.R., Bell J.C., Harwood C.E., Owen J.V. (1996) The effect of

outcrossing rate on the growth of selected families of *Eucalyptus grandis*. *Silvae Genetica* 45, 97-101.

Butcher P.A., Williams E.R. (2002) Variation in outcrossing rates and growth in *Eucalyptus camaldulensis* from the Petford Region, Queensland; Evidence of outbreeding depression. *Silvae Genetica* 51, 6-12.

Butcher T.B. (1990) '*Eucalyptus globulus* tree improvement cooperative in Western Australia.' (Department of Conservation and Land Management: Perth, Western Australia)

Byrne M., Parrish T.L., Moran G.F. (1998) Nuclear RFLP diversity in *Eucalyptus nitens*. *Heredity* 81, 225-233.

Cañas I., Soria F., López G., Astorga R., Toval G. (2004) Genetic parameters for rooting in *Eucalyptus globulus* (Labill.). In proceedings of IUFRO Conference 'Eucalyptus in a changing world'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 159-160. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Carnegie A.J., Keane P.J., Ades P.K., Smith I.W. (1994) Variation in susceptibility of *Eucalyptus globulus* provenances to *Mycosphaerella* leaf disease. *Canadian Journal of Forest Research* 24, 1751-1757.

Chalmers P. (1992) The adaptive significance of juvenile versus adult leaves in *Eucalyptus globulus* ssp. *globulus*. B.Sc. (Hons.) thesis, University of Tasmania.

Chambers P.G.S., Borralho N.M.G., Potts B.M. (1996) Genetic analysis of survival in *Eucalyptus globulus* ssp. *globulus*. *Silvae Genetica* 45, 107-112.

Clarke F.R., Baker R.J. (1996) Spatial analysis improves comparison of seed lot comparisons. *Crop Science* 36, 1180-1184.

Cook I.O., Ladiges P.Y. (1991) Morphological variation within *Eucalyptus nitens* s. lat. and recognition of a new species, *E. denticulata*. *Australian Systematic Botany* 4, 375-390.

Cook I.O., Ladiges P.Y. (1998) Isozyme variation in *Eucalyptus nitens* and *E. denticulata*. *Australian Journal of Botany* 46, 35-44.

Correll R.L., Cellier K.M. (1987) Effects of plot size, block size and buffer rows on the precision of forestry trials. *Australian Forest Research* 17, 11-18.

Costa e Silva J., Dutkowski G.W., Borralho N.M.G. (2005) Across-site heterogeneity of genetic and environmental variances in the genetic evaluation of Eucalyptus globulus trials for height growth. *Annals of Forest Science* 62, 183-191.

Costa e Silva J., Dutkowski G.W., Gilmour A.R. (2001) Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Canadian Journal of Forest Research* 31, 1887-1893.

Cremer K.W. (1977) Distance of seed dispersal in eucalypts estimated from seed weights. *Australian Forest Research* 7, 225-228.

Cullis B., Gleeson A.C. (1989) Efficiency of neighbour analysis for replicated field trials in Australia. *Journal of Agricultural Science* 113, 233-239.

Cullis B., Gogel B., Verbyla A., Thompson R. (1998) Spatial analysis of multi-environment early generation variety trials. *Biometrics* 54, 1-18.

Cullis B.R., A.C.Gleeson, Thomson F.M. (1992) The response to selection of different procedures for the analysis of early generation variety trials. *Journal of Agricultural Science, Cambridge* 118, 141-148.

Cullis B.R., Gleeson A.C. (1991) Spatial analysis of field experiments - an extension to two dimensions. *Biometrics* 47, 1449-1460.

Cullis B.R., Lill W., J., Fisher J.A., Read B.J., Gleeson A.C. (1989) A new procedure for the analysis of early generation variety trials. *Applied Statistics* 38, 361-375.

de Resende M.D.V., Stringer J., Cullis B., Thompson R. (2004) Joint modelling of competition and spatial variability in forest field trials. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 330-332. (RAIZ, Instituto Inverstigação da Floresta e Papel: Aveiro, Portugal)

de Resende M.D.V., Thompson R. (2004) Factor analytic multiplicative mixed models in the analysis of multi-environment forest field trials. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 123-129. (RAIZ, Instituto Inverstigação da Floresta e Papel: Aveiro, Portugal)

De Veer C., Balocchi C., Crespell P., Hodge G., Droppelmann F. (2001) Biocat: un programa para el análisis de ensayos genéticos forestales. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15th September, Valdivia, Chile. (INFOR: Santiago, Chile)

de-Souza E.A., Geraldi I.O., Ramalho M.A.P., Bertolucci F.d.L.G. (2003) Experimental alternatives for evaluation of progenies and clones in eucalyptus breeding programs. *Revista Árvore* 27, 427-434.

Diggle P.J., Liang K.-Y., Zeger S.L. (1994) 'Analysis of longitudinal data.' (Clarendon Press: Oxford)

Durban M., Currie I.D., Kempton R.A. (2001) Adjusting for fertility and competition in variety trials. *Journal of Agricultural Science* 136, 129-140.

Dutkowski G.W. (1995) Genetic variation in drought susceptibility of *Eucalyptus globulus* ssp. *globulus* in plantations in Western Australia. In proceedings of CRCTHF-IUFRO Conference 'Eucalypt Plantations: Improving Fibre Yield and Quality'. (Eds B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits

and C.A. Raymond). 19-24 February, Hobart, Tasmania. pp. 199-203. (CRC for Temperate Hardwood Forestry: Hobart, Tasmania)

Dutkowski G.W. (2000a) '*Eucalyptus globulus* race classification confirmed and extended.' Hot Off the Seed Bed 32, 2 pp. (CRC-SPF: Hobart, Tasmania)

Dutkowski G.W. (2000b) 'Prediction of breeding values for diameter and pilodyn penetration for Fletcher Challenge Paper *Eucalyptus nitens* breeding population families.' Confidential Consultancy Report, 8 pp. (CRCSPF: Hobart, Tasmania)

Dutkowski G.W. (2004) Gains from clonal breeding and deployment. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. p. 163. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Dutkowski G.W., Costa e Silva J., Gilmour A.R., Lopez G.A. (2002) Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research* 32, 2201-2214.

Dutkowski G.W., Potts B.M. (1999) Geographical patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Australian Journal of Botany* 46, 237-263.

Dutkowski G.W., Potts B.M., Borralho N.M.G. (1997) Revised racial classification of *Eucalyptus globulus* ssp. *globulus* and the importance of including race in analysis of progeny trials. In proceedings of 'IUFRO Conference on Silviculture and Improvement of Eucalypts'. (Eds A.R. Higa, E. Schaitza and S. Gaiad). 24-29 August, Salvador, Bahia, Brazil. Vol. 1. pp. 322-329. (EMBRAPA: Colombo, Brazil)

Dutkowski G.W., Raymond C.A. (2001) A decision support tool for expensive to measure traits in progeny trials. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. (INFOR: Santiago, Chile)

Eldridge K., Davidson J., Harwood C., van Wyk G. (1993) 'Eucalypt domestication and breeding.' (Clarendon Press: Oxford)

Endler J.A. (1977) 'Geographical variation, speciation and clines.' (Princeton: Princeton University Press)

Ericsson T. (1997) Enhanced heritabilities and Best Linear Unbiased Predictors through appropriate blocking of progeny trials. *Canadian Journal of Forest Research* 27, 2097-2101.

Fagroud M., Van Meirvenne M. (2002) Accounting for soil spatial autocorrelation in the design of experimental trials. *Soil Science Society of America Journal* 66, 1134-1142.

Falconer D.S., MacKay T.F.C. (1996) 'Introduction to Quantitative Genetics.' (Longman Scientific and Technical: New York)

Famula T.R. (1992) Simple and rapid inversion of additive relationship matrices incorporating parental uncertainty. *Journal of Animal Science* 70, 1045-1048.

Farrow R.A., Floyd R.B., Newmann F.G. (1994) Inter-provenance variation in resistance of *Eucalyptus globulus* to insect feeding. *Australian Forestry* 57, 65-68.

Federer W.T. (1998) Recovery of interblock, intergradient and intervariety information in incomplete block and lattice rectangle designed experiments. *Biometrics* 54, 471-481.

Fernandez A., Rodriguez M.C., Ipinza R. (1998) Estimacion bayesiana de parametros geneticos y efectos de procedencia de caracteres de crecimiento en *Eucalyptus nitens*. *ITEA Produccion Animal* 94A: 3, 279-289.

Fox J.C., Ades P.K., Bi H. (2001) Stochastic structure and individual-tree growth models. *Forest Ecology and Management* 154, 261-276.

Freeman J.S., Jackson H.D., Steane D.A., McKinnon G.E., Dutkowski G.W., Potts B.M., Vaillancourt R.E. (2001) Chloroplast DNA phylogeography of *Eucalyptus globulus*. *Australian Journal of Botany* 49, 585-596.

Fu Y. (2003) On implementation of incomplete block designs in forest genetic field trials. *Forest Genetics* 10, 23-33.

Fu Y.B., Yanchuk A.D., Namkoong G. (1999) Spatial patterns of tree height variations in a series of Douglas-fir progeny trials: implications for genetic testing. *Canadian Journal of Forest Research* 29, 714-723.

Gaiotto F.A., Bramucci, M. and Grattapaglia, D (1997) Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* with dominant RAPD and AFLP markers. In proceedings of 'IUFRO Conference on Silviculture and Improvement of Eucalypts'. (Eds A.R. Higa, E. Schaitza and S. Gaiad). 24-29 August, Salvador, Bahia, Brazil. Vol. 1. pp. 53-59. (EMBRAPA: Colombo, Brazil)

Gardiner C.A., Crawford D.A. (1987) '1987 Seed Collections of *Eucalyptus globulus* subsp. *globulus* for tree improvement purposes.' (CSIRO Division of Forest Research: Canberra, Australia)

Gardiner C.A., Crawford D.A. (1988) '1988 Seed Collections of *Eucalyptus globulus* subsp. *globulus* for tree improvement purposes.' (CSIRO Division of Forest Research: Canberra, Australia)

Gilmour A.R., Cullis B.R., Verbyla A.P. (1997a) Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological and Environmental Statistics* 2, 269-293.

Gilmour A.R., Cullis B.R., Welham S.J., Thompson R. (1999) 'ASREML.' Technical Report. (NSW Agriculture: Orange)

Gilmour A.R., Thompson R., Cullis B.R., Welham S.J. (1997b) 'ASREML.' Technical

Report. (NSW Agriculture: Orange)

Gleeson A.C., Cullis B.R. (1987) Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* 43, 277-288.

Greaves B.L., Borralho N.M.G., Raymond C.A., Farrington A. (1996) Use of a pilodyn for the indirect selection of basic density in *Eucalyptus nitens*. *Canadian Journal of Forest Research* 26, 1643-1650.

Greaves B.L., Hamilton M., Pilbeam D.J., Dutkowski G.W. (2004) Genetic variation in commercial properties of six and fifteen year-old *Eucalyptus globulus*. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 97-102. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Groeneveld E. (1990) 'PEST User's Manual.' (Institute of Animal Husbandry and Animal Behaviour, Federal Research Centre of Agriculture, Germany)

Groeneveld E. (1996) 'REML VCE - A Multivariate Multi Model Restricted Maximum Likelihood (Co)Variance Component Estimation Package.' (Institute of Animal Husbandry and Animal Ethology, Federal Research Centre of Agriculture, Germany)

Grondona M.O., Crossa J., Fox P.N., Pfeiffer W.H. (1996) Analysis of variety yield trials using two-dimensional separable ARIMA processes. *Biometrics* 52, 763-770.

Guimaraes M.P., Almeida M.H., Tome M., Pereira H. (1995) Variation in tree taper and DBH relationships among provenances of *Eucalyptus globulus*. In proceedings of CRCTHF-IUFRO Conference 'Eucalypt Plantations: Improving Fibre Yield and Quality'. (Eds B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits and C.A. Raymond). 19-24 February, Hobart, Tasmania. pp. 371-372. (CRC for Temperate Hardwood Forestry: Hobart, Tasmania)

Hamann A., Kosky M.P., Namkoong G. (2001) Improving precision of breeding values by removing spatially autocorrelated variation in forestry field experiments. *Silvae Genetica* 51, 210-215.

Hardner C., Borralho N., Tier B., Miller S., Goddard M. (1996a) Accounting for dominance and inbreeding in genetic evaluations using individual tree mixed models. In proceedings of QFRI - IUFRO Conference 'Tree Improvement for Sustainable Tropical Forestry'. (Eds M.J. Dieters, A.C. Matheson, D.G. Nikles, C.E. Harwood and S.M. Walker). 27 October - 1 November, Caloundra, Queensland. Vol. 1. pp. 143-147. (Queensland Forest Research Institute: Gympie, Queensland)

Hardner C.M., Potts B.M. (1995) Inbreeding depression and changes in variation after selfing *Eucalyptus globulus* subsp. *globulus*. *Silvae Genetica* 44, 46-54.

Hardner C.M., Potts B.M., Gore P.L. (1998) The relationship between cross success and spatial proximity of *Eucalyptus globulus* ssp. *globulus* parents. *Evolution* 52, 614-

618.

Hardner C.M., Vaillancourt R.E., Potts B.M. (1996b) Stand density influences outcrossing rate and growth of open-pollinated families of *Eucalyptus globulus*. *Silvae Genetica* 45, 226-228.

Hasan O. (1993) The effects of paclobutrazol on flowering activity and gibberellin levels in *Eucalyptus nitens* and *E. globulus*. Ph. D. thesis, University of Tasmania.

Henderson C.R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 32, 69.

Henderson C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics* 32, 69-83.

Henderson C.R. (1984) 'Applications of Linear Models in Animal Breeding.' (University of Guelph: Guelph)

Henderson C.R. (1987) Use of an average numerator relationship matrix for multiple-sire joinings. *Journal of Animal Science* 66, 1614-1621.

Hodge G.R., Volker P.W., Potts B.M., Owen J.V. (1996) A comparison of genetic information from open-pollinated and control-pollinated progeny tests in two eucalypt species. *Theoretical and Applied Genetics* 92, 53-63.

Huhn M., Langner W. (1995) Investigations on the correlation pattern in even-aged stands of larch. II. Dynamic description of phenotypic correlations between neighbouring observations. *Silvae Genetica* 44, 325-338.

Ipinza R., Gutiérrez B. (1999) Progeny and provenance tests in Chile. Evaluation by spatial analysis (ASREML) in *Nothofagus alpina,*a case study. (Ensayos de orogenie y procedencia en Chile. Alálysis espacial (ASREML) en *Nothofagus alpina*, un estudio de caso.). In proceedings of IUFRO International Symposium 'Long-term observations and research in forestry'. (Eds C. Kleinn and M. Köhl). 23-27 February, Turrialba, Costa Rica. Vol. 2. pp. 187-203. (Catie)

Isaaks E.H., Srivastava R.M. (1989) 'An Introduction to Applied Geostatistics.' (Oxford University Press: New York)

Jarvis S.F., Borralho N.M.G., Potts B.M. (1995) Implementation of a multivariate BLUP model for genetic evaluation of *Eucalyptus globulus* in Australia. In proceedings of CRCTHF-IUFRO Conference 'Eucalypt Plantations: Improving Fibre Yield and Quality'. (Eds B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits and C.A. Raymond). 19-24 February, Hobart, Tasmania. pp. 212-216. (CRC for Temperate Hardwood Forestry: Hobart, Tasmania)

Jennrich R.I., Schluchter M.D. (1986) Unbalanced Repeated-Measures Models with Structured Covariance Matrices. *Biometrics* 42, 805-820.

Jones R.C., Steane D.A., Potts B.M., Vaillancourt R.E. (2002) Microsatellite and morphological analysis of *Eucalyptus globulus* populations. *Canadian Journal of*

Forest Research 32, 59-66.

Jones R.C., Steane D.A., Vaillancourt R.E., Potts B.M. (2001) Molecular affinities of the Jeeralang population of *Euclayptus globulus*. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. (INFOR: Santiago, Chile)

Jordan G.J., Borralho N.M.G., Tilyard P., Potts B.M. (1994) Identification of races in *Eucalyptus globulus* spp *globulus* based on growth traits in Tasmania and geographic distribution. *Silvae Genetica* 43, 292-298.

Jordan G.J., Dutkowski G.W., Potts B.M., MacDonald A.C., Tilyard P., Borralho. N.M.G. (1998) 'Genetic variation in North Forest Products' *Eucalyptus globulus* ssp. *globulus* base population trials.' Technical Report 8. (Cooperative Research Centre for Sustainable Production Forestry: Hobart, Tasmania, Australia)

Jordan G.J., Potts B.M., Kirkpatrick J.B., Gardiner C. (1993) Variation in the *Eucalyptus globulus* complex revisited. *Australian Journal of Botany* 41, 763-85.

Joyce D., Ford R., Fu Y.B. (2002) Spatial patterns of tree height variations in a black spruce farm-field progeny test and neighbors-adjusted estimations of genetic parameters. *Silvae Genetica* 51, 13-18.

Kempton R.A., Seraphin J.C., Sword A.M. (1994) Statistical analysis of two-dimensional variation in variety yield trials. *Journal of Agricultural Science* 111, 339-346.

Kennedy B.W. (1989) 'Animal Model BLUP, Erasmus Intensive Graduate Courese.' (Centre for Genetic Improvement of Livestock, University of Guelph)

Kenward M.G., Roger J.H. (1997) Small sample inference for fixed effetcs from restricted maximum likelihood. *Biometrics* 53, 983-997.

Kerr R.J., McRae T.A., Dutkowski G.W., Apiolaza L.A., Tier B. (2001) TREEPLAN - a genetic evaluation system for forest tree improvement. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. (INFOR: Santiago, Chile)

Kirk H.J., Haynes F.L., Monroe R.J. (1980) Application of trend analysis to horticultural field trials. *Journal of the American Society of Horticultural Science* 105, 189-193.

Kirkpatrick J.B. (1975) 'Geographical Variation in *Eucalyptus globulus*.' (Australian Government Publishing Service: Canberra)

Kube P.D., Dutkowski G.W. (2002) 'Estimation of *Eucalyptus nitens* breeding values for volume, basic density, pulp yield and branch size at RP252/8 & 9.' Technical Report 28/2002, 21 pp. (Forestry Tasmania: Hobart, Tasmania)

Kube P.D., Matheson A.C., Raymond C.A., Boomsma D.B. (1995) Provenance variation in *Eucalyptus globulus*. In proceedings of CRCTHF-IUFRO Conference

'Eucalypt Plantations: Improving Fibre Yield and Quality'. (Eds B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits and C.A. Raymond). 19-24 February, Hobart, Tasmania. pp. 245-246. (CRC for Temperate Hardwood Forestry: Hobart, Tasmania)

Kusnandar D., Galwey N. (2000) A proposed method for estimation of genetic parameters on forest trees without raising progeny: Critical evaluation and refinement. *Silvae Genetica* 49, 15-21.

Latta R.G. (2004) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist* 161, 51-58.

Libby W.J., Cockerham C.C. (1980) Random non-contiguous plots in interlocking field layouts. *Silvae Genetica* 29, 183-190.

Lill W.J., Gleeson A.C., Cullis B.R. (1988) Relative accuracy of a neighbour method for field trials. *Journal of Agricultural Science Cambridge* 111, 339-346.

Liu J.P., Burkhart H.E. (1994) Spatial characteristics of diameter and total height in juvenile Loblolly Pine (*Pinus taeda* L) plantations. *Forest Science* 40, 774-786.

Loo-Dinkins J.A., Tauer C.G. (1987) Statistical efficiency of six progeny test field designs on three loblolly pine (*Pinus taeda* L.) site types. *Canadian Journal of Forest Research* 17, 1066-1070.

Lopez G.A., Potts B.M., Dutkowski G.W., Apiolaza L.A., Gelid P. (2002) Genetic variation and inter-trait correlations in *Eucalyptus globulus* base population trials in Argentina. *Forest Genetics* 9, 223-237.

Lopez G.A., Potts B.M., Dutkowski G.W., Traverso J.M.R. (2001) Quantitative genetics of *Eucalyptus globulus*: Affinities of land race and native stand localities. *Silvae Genetica* 50, 244-252.

Lorkin M.H., Potts B.M., Vaillancourt R.E., Williams D.R., Gore P.L. (2004) 'Vulnerability of *Eucalyptus globulus* to the seed destroying wasp *Megastigmus sp*. is under genetic control.' Hot Off the Seed Bed 69, 2 pp. (CRC-SPF: Hobart, Tasmania)

MacDonald A.C., Borralho N.M.G., Potts B.M. (1998) Genetic variation for growth and wood density in *Eucalyptus globulus* ssp. *globulus* in Tasmania (Australia). *Silvae Genetica* 46, 236-241.

MacDonald A.C., Borralho N.M.G., Potts B.M., Jarvis S.F. (1995) International analysis of a comprehensive *Eucalyptus globulus* ssp. Labill. *globulus* base population collection. In proceedings of CRCTHF-IUFRO Conference 'Eucalypt Plantations: Improving Fibre Yield and Quality'. (Eds B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits and C.A. Raymond). 19-24 February, Hobart, Tasmania. pp. 247-248. (CRC for Temperate Hardwood Forestry: Hobart, Tasmania)

Magnussen S. (1989) Effects and adjustments of competition bias in progeny trials with single-tree plots. *Forest Science* 35, 532-547.

Magnussen S. (1990) Application and comparison of spatial models in analyzing tree-genetics field trials. *Canadian Journal of Forest Research* 20, 536-546.

Magnussen S. (1993) Bias in genetic variance due to spatial autocorrelation. *Theoretical and Applied Genetics* 86, 349-355.

Magnussen S. (1994) A method to adjust simultaneously for spatial microsite and competition effects. *Canadian Journal of Forest Research* 24, 985-995.

McGowen M.H., Potts B.M., Vaillancourt R.E., Gore P., Williams D.R., Pilbeam D.J. (2004) The genetic control of sexual reproduction in *Eucalyptus globulus*. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 104-108. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

McRae T.A., Dutkowski G.W., Pilbeam D.J., Powell M.B., Tier B. (2004) Genetic Evaluation Using the TREEPLAN® System. In proceedings of IUFRO Joint Conference Division 2 'Forest Genetics and Tree Breeding in the Age of Genomics: Progress and Future'. (Eds B. Li and S. McKeand). 1-5 November, Charleston, South Carolina. (North Carolina State University: Raleigh, North Carolina)

Mora-Garces A., Ramirez G. (2000) Practical consideration on the spatial analysis of forest field trials (Consideraciones practicas sobre el analisis estadistico espacial de experimentos forestales de campo). *Revista Forestal Venezolana* 44, 21-28.

Moreau L., Monod H., Charcosset A., Gallais A. (1999) Marker-assisted selection with spatial analysis of unreplicated field trials. *Theoretical & Applied Genetics* 98, 234-242.

Mrode R.A. (1998) 'Linear Models for the Prediction of Breeding Values.' (CAB International: Wallingford, U.K.)

Mullin T.J., Park Y.S. (1995) Stochastic simulation of population management strategies for tree breeding - a new decision-support tool for personal computers. *Silvae Genetica* 44, 132-141.

Nelder J.A. (1965) The analysis of randomised experiments with orthogonal block structure. II Block structure and the null analysis of variance. *Proceedings of the Royal Society of London, Series A* 283, 147-162.

Nesbitt K.A., Potts B.M., Vaillancourt R.E., West A.K., Reid J.B. (1995) Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). *Heredity* 74, 628-637.

Nguyen N., Williams E.R. (1993) An algorithm for constructing optimal resolvable row-column designs. *Australian Journal of Statistics* 35, 363-370.

O'Reilly-Wapstra J.M., McArthur C., Potts B.M. (2001) Genetic variation in resistance of *Eucalyptus globulus* to marsupial browsers. *Oecologia* 130, 289-296.

Orme R.K. (1977) *E. globulus* provenances. In proceedings of 'Third World Consultation on Forest Tree Breeding'. 21-26 March, Canberra, Australia. pp. 207-221.

Papadakis J.S. (1937) 'Méthode statistique pour des expériences sur champ.' Bulletin 23. (Institute de Amélioration des Plantes à Salonique)

Patterson B., Potts B.M., Vaillancourt R.E. (2001) Variation in outcrossing rates within the canopy of *Eucalyptus globulus*. In proceedings of IUFRO International Symposium 'Developing the Eucalypt of the Future'. 10-15 September, Valdivia, Chile. p. 132. (INFOR: Santiago, Chile)

Patterson H.D., Thompson R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545-554.

Patterson H.D., Williams E.R. (1976) A new class of resolvable incomplete block designs. *Biometrika* 63, 83-92.

Payne R.W., Lane P.W.*, et al.* (1988) 'GENSTAT 5 Reference Manual.' (Clarendon Press: Oxford)

Pederick L.A. (1979) Natural variation in Shining Gum (*Eucalyptus nitens*). *Australian Forest Research* 9, 41-63.

Pederick L.A., Hopmans P., Flinn D., W., Abbott I.D. (1984) Variation in genetic response to suspected copper deficiency in *Pinus radiata*. *Australian Forest Research* 14, 75-84.

Perez-Enciso M., Fernando R.L. (1992) Genetic evaluation of uncertain parentage: a comparison of methods. *Theoretical and Applied Genetics* 84, 173-179.

Piepho H.-P. (1998) Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics* 97, 195-201.

Pilbeam D.J., Dutkowski G.W. (2004) Simulation to determine optimal wood quality sampling strategies. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 211-212. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Potts B.M., Jordan G.J. (1994a) Genetic variation in the juvenile leaf morphology of *Eucalyptus globulus* Labill. ssp. *globulus*. *Forest Genetics* 1, 81-95.

Potts B.M., Jordan G.J. (1994b) The spatial pattern and scale of variation in *Eucalyptus globulus* ssp *globulus*: variation in seedling abnormalities and early growth. *Australian Journal of Botany* 42, 471-492.

Potts B.M., Milgate A., Joyce K., Mohammed C.L., Vaillancourt R.E., Dutkowski G.W. (2004a) Quantitative genetic control of *Mycosphaerella* resistance in *Eucalyptus globulus* and impact on growth. In proceedings of IUFRO Conference

'Eucalyptus in a changing world'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 116-122. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Potts B.M., Vaillancourt R.E., *et al.* (2004b) Exploration of the *Eucalyptus globulus* gene pool. In proceedings of IUFRO Conference 'Eucalyptus in a changing world'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 46-61. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Prado J.A., Alvear C. (1993) Resultados de ensayos de procedencias y progenies de *Eucalyptus globulus* en la zona centro-sur de chile. In proceedings of 'Los Eucalyptus en el desarrollo Forestal de Chile'. (Eds A.S. Barros, J.A. Prado and C. Alvear) Chile. pp. 61-79. (Instituto Forestal Filial Corfo and Corporacion de Fomento de la Produccion)

Qiao C., Basford K., DeLacy I., Cooper M. (2003) Advantage of single-trial models for response to selection in wheat breeding multi-environment trials. *Theoretical and Applied Genetics* 106, 1541-1559.

Qiao C.G., Basford K.E., DeLacy I.H., Cooper M. (2000) Evaluation of experimental designs and spatial analyses in wheat breeding trials. *Theoretical & Applied Genetics* 100, 9-16.

Quaas R.L. (1978) Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32, 949-953.

Quaas R.L. (1981) Additive genetic model with groups and relationships. *Journal of Dairy Science* 71, 1338-1345.

Quaas R.L., Anderson R.D., Gilmour A.R. (1984) 'BLUP School Handbook. Use of Mixed Models for Prediction and for Estimation of (Co)Variance Components.' (Animal Genetics and Breeding Unit: Armidale, N.S.W.)

Rapley L.P., Allen G.R., Potts B.M. (2004a) Genetic variation in *Eucalyptus globulus* in relation to susceptibility from attack by the southern eucalypt leaf beetle, *Chrysophtharta agricola*. *Australian Journal of Botany* 52, 747–756.

Rapley L.P., Allen G.R., Potts B.M. (2004b) Genetic variation of *Eucalyptus globulus* in relation to autumn gum moth *Mnesampela privata* (Lepidoptera: Geometridae) oviposition preference. *Forest Ecology and Management* 194, 169-175.

Raymond C.A., Schimleck L.R., Muneri A., Michell A.J. (2001) Genetic parameters and genotype by environment interactions for pulp yield predicted using near infrared reflectance analysis and pulp productivity in *Eucalyptus globulus*. *Forest Genetics* 8, 213-224.

Reed D.D., Burkhart H.E. (1985) Spatial autocorrelation of individual tree characteristics in loblolly pine stands. *Forest Science* 31, 575-587.

Roughgarden J. (1979) 'Theory of Population Genetics and Evolutionary Ecology: An

Introduction.' (Macmillan Publishing: New York)

Saenz-Romero C., Nordheim E.V., Guries R.P., Crump P.M. (2001) A case study of provenance/progeny testing using trend analysis with correlated errors and SAS PROC MIXED. *Silvae Genetica* 50, 127-135.

SAS Institute (1990) 'SAS/STAT User's Guide, Version 6 (Fourth Edition).' (SAS Institute: Cary, NC)

Singh M., Malhotra R.S., Ceccarelli S., Sarker A., Grando S., Erskine W. (2003) Spatial variability models to improve dryland field trials. *Experimental Agriculture* 39, 151-160.

Skabo S., Vaillancourt R.E., Potts B.M. (1998) Fine-scale genetic structure of *Eucalyptus globulus* ssp. *globulus* forest revealed by RAPDs. *Australian Journal of Botany* 46, 583-594.

Smith K.F., Casler M.D. (2004) Spatial analysis of forage grass trials across locations, years, and harvests. *Crop Science* 44, 56-62.

Sneath P.H.A., Sokal R.R. (1973) 'Numerical Taxonomy.' (W.H. Freeman and Co.: San Francisco)

Snedecor G.W., Cochrane W.G. (1980) 'Statistical Methods.' (The Iowa State University Press: Ames, Iowa)

Sokal R.R., Rohlf F.L. (1981) 'Biometry.' (W. H. Freeman: San Francisco)

Soria F., Basurco F., Toval G., Silio L., Rodriguez M.C., Toro M. (1998) An application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. *Canadian Journal of Forest Research* 28, 1286-1294.

Soria F., Borralho N.M.G. (1998) The genetics of resistance to *Phoracantha semipunctata* attack in *Eucalyptus globulus* in Spain. *Silvae Genetica* 46, 365-369.

Spencer D.J., Williams E.R. (1993) Effluent-irrigated *Eucalyptus globulus* progeny trial: first results. In proceedings of 'Meeting of Research Working Group 1, Forest Genetics' Coonawarra, South Australia. (Australian Forestry Council: Canberra, Australia)

Squillace A.E. (1974) Average genetic correlations among offspring from open-pollinated forest trees. *Silvae Genetica* 23, 149-156.

Steane D.A., Vaillancourt R.E., Russell J., Powell W., Marshall D., Potts B.M. (2001) Microsatellite variation in a forest tree, *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). *Silvae Genetica* 50, 89-91.

Stram D.O., Lee J.W. (1994) Variance components testing in the longitudinal mixed effects model. *Biometrics* 50, 1171-1177.

Stringer J.K., Cullis B.R. (2002) Application of spatial analysis techniques to adjust

for fertility trends and identify interplot competition in early stage sugarcane selection trials. *Australian Journal of Agricultural Research* 53, 911-918.

Tamura R.N., Nelson L.A., Naderman G.C. (1988) An investigation of the validity and usefuleness of trend surface analysis for field plot data. *Agronomy Journal* 80, 712-718.

Tibbits W., Hodge G. (1998) Genetic parameters and breeding value predictions for *Eucalyptus nitens* wood fiber production traits. *Forest Science* 44, 587-598.

Tibbits W.N., Boomsma D.B., Jarvis S. (1997) Distribution, biology, genetics, and improvement programs for *Eucalyptus globulus* and *E. nitens* around the world. In proceedings of '24th Southern Forest Tree Improvement Conference'. (Eds T. White, D. Huber and G. Powell). 9-12 June, Orlando, Florida. pp. 1-15. (Southern Tree Improvement Committee)

Tibbits W.N., Hodge G.R., White T.L. (1991) Predicting breeding values for freezing resistance in *Eucalyptus globulus*. In proceedings of IUFRO Symposium 'Intensive Forestry: The Role of Eucalypts'. (Ed. A.P.G. Schonau) Durban, South Africa. Vol. 1. pp. 330-333. (Southern African Institute of Forestry)

Toro M.A., Silió L., Rodriguez M.C., Soria F., Toval G. (1998) Genetic analysis of survival to drought in *Eucalyptus globulus* in Spain. In proceedings of '6th World Congress on Genetics Applied to Livestock Production'. January, Armidale, Australia. Vol. 27. pp. 499-502. (Organising Committee, 6th World Congress on Genetics Applied to Livestock Production: Armidale, Australia)

Toval G. (2004) The *Eucalyptus globulus* clonal silviculture in mediterranean climate. In proceedings of IUFRO Conference 'Eucalyptus in a Changing World'. (Eds N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira and M. Tomé). 11-15 October, Aveiro, Portugal. pp. 70-78. (RAIZ, Instituto Investigação de Floresta e Papel: Aveiro, Portugal)

Vega Alonso G., Toval Hernandez G., Basurco Garcia-Casal F., Rodriguez Soalleiro R. (1994) Early growth of 260 families of *Eucalyptus globulus* in North Spain. In proceedings of '*Eucalyptus* for Biomass Production'. (Eds J. Pereira and H. Pereira) Lisbon, Portugal. pp. 115-132. (Commission of the European Communities)

Verbyla A.P., Cullis B.R., Kenward M.G., Welham S.J. (1997) 'The analysis of designed experiments and longitudinal data using smoothing splines.' Research Report 97/4. (The University of Adelaide: Adelaide, Australia)

Volker P.W. (2002) Quantitative genetics of *Eucalyptus globulus*, *E. nitens* and their F1 hybrid. Ph.D. thesis, University of Tasmania.

Volker P.W., Orme R.K. (1988) Provenance trials of *Eucalyptus globulus* and related species in Tasmania. *Australian Forestry* 51, 257-265.

Wei X., Borralho N.M.G. (1998) Use of individual tree mixed models to account for mortality and selective thinning when estimating base population genetic parameters.

Forest Science 44, 246-253.

Wei X.M., Borralho M.G. (2000) Genetic gains and levels of relatedness from best linear unbiased prediction selection of *Eucalyptus urophylla* for pulp production in southeastern China. *Canadian Journal of Forest Research* 30, 1601-1607.

Westell R.A., Quaas R.L., van Vleck L.D. (1988) Genetic groups in an animal model. *Journal of Dairy Science* 71, 1310-1318.

White T.L., Hodge G.R. (1989) 'Predicting Breeding Values With Applications in Forest Tree Improvement.' (Kluwer Academic Publishers: London, Dordrecht)

Whittock S.P., Apiolaza L.A., Kelly C.M., Potts B.M. (2003) Genetic control of coppice and lignotuber development in *Eucalyptus globulus*. *Australian Journal of Botany* 51, 57-67.

Wilkinson G.N., Eckert S.R., Hancock T.W., Mayo O. (1983) Nearest neighbour (NN) analysis of field experiments. *Journal of the Royal Statistical Society* 45, 151-211.

Williams E.R. (1986) A neighbour model for field experiments. *Biometrika* 73, 279-287.

Williams E.R., John J.A., Whitaker D. (2005) Construction of resolvable spatial row-column designs. *Biometrics* In press.

Williams E.R., Matheson A.C. (1994) 'Experimental Design and Analysis for Use in Tree Improvement.' (CSIRO: Melbourne)

Wright S. (1934) The method of path coefficients. *Annals of Mathematical Statistics* 5, 161-215.

Yang R., Ye T., Blade S., Bandara M. (2004) Efficiency of spatial analyses of field pea variety trials. *Crop Science* 44, 49-55.

Ye T.Z., Jayawickrama K.J.S., Johnson G.B. (2004) Efficiency of using first-generation information during second-generation selection: results of computer simulation. In proceedings of IUFRO Conference Division 2 'Forest Genetics and Tree Breeding in the Age of Genomics: Progress and Future'. (Eds B. Li and S. McKeand) Charleston, South Carolina. pp. 214-217. (North Carolina State University)

Zang D., Wang H., You Y. (1995) Performance and selection of a 4-year *Eucalyptus globulus* seedling seed orchard in Yunnan, China. In proceedings of CRCTHF-IUFRO Conference 'Eucalypt Plantations: Improving Fibre Yield and Quality'. (Eds B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits and C.A. Raymond). 19-24 February, Hobart, Tasmania. pp. 226-229. (CRC for Temperate Hardwood Forestry: Hobart, Tasmania)

Zimmerman D.L., Harville D.A. (1991) A random field approach to the analysis of

field-plot experiments and other spatial experiments. *Biometrics* 47, 223-239.